



Identification d'espèces végétales par une description géométrique locale d'images de feuilles

Sofiène Mouine

► To cite this version:

Sofiène Mouine. Identification d'espèces végétales par une description géométrique locale d'images de feuilles. Traitement du signal et de l'image [eess.SP]. Télécom ParisTech, 2015. Français. NNT : 2015ENST0016 . tel-01223709v2

HAL Id: tel-01223709

<https://inria.hal.science/tel-01223709v2>

Submitted on 17 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



2015-ENST-0016

Doctorat ParisTech

T H È S E

présentée pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Signal et Image »

présentée et soutenue publiquement le 2 avril 2015 par
Sofiene MOUINE

Identification d'espèces végétales par une description géométrique locale d'images de feuilles

JURY

Valérie GOUET-BRUNET, Directrice de recherche au laboratoire MATIS, IGN

Laure TOUGNE, Professeur au LIRIS, Université Lyon 2

Mounim A. EL YACOUBI, Professeur à Intermedia, Télécom SudParis

Laurent WENDLING, Professeur au LIPADE, Université Paris Descartes

Itheri YAHIAOUI, Maître de conférence au CReSTIC, Université de Reims

Anne VERROUST-BLONDET, Chargée de recherche à l'Inria Paris-Rocquencourt

Rapporteur

Rapporteur

Examineur

Examineur

Co-directeur de thèse

Directeur de thèse

Remerciements

Beaucoup de personnes ont contribué de près ou de loin à l'aboutissement de ce travail de thèse.

Je remercie tout d'abord Anne Verroust-Blondet pour avoir accepté de diriger cette thèse. Grâce à sa rigueur scientifique, j'ai beaucoup appris durant ces années de thèse. J'ai aussi apprécié ses précieux conseils qui m'ont guidé dans la réflexion et la rédaction scientifique.

J'adresse un grand merci à ma co-encadrante Itheri Yahiaoui. Ses qualités scientifiques et humaines m'ont beaucoup aidé durant cette thèse. J'ai apprécié son implication totale durant toute la thèse. Ses encouragements incessants et son soutien ont permis à ce travail d'aboutir. Je la remercie également d'avoir été toujours à l'écoute dans les moments difficiles.

Je remercie les membres du jury Mme. Laure Tougne, Mme. Valérie Gouet-Brunet, M. Laurent Wendling et M. Mounim A. El Yacoubi pour avoir accepté de faire partie du jury de cette thèse et pour le temps consacré à la lecture du manuscrit.

Je remercie Nozha Boujemaa d'avoir retenu ma candidature pour effectuer un stage puis une thèse dans le cadre du projet Pl@ntNet, me permettant ainsi d'avoir une première expérience dans le monde de la recherche scientifique. Je remercie aussi tous les botanistes du projet Pl@ntNet avec qui j'ai pu échanger. J'ai grandement profité de leur expertise en botanique. Nos échanges ont été fructueux et essentiels pour le bon déroulement de ce travail. Je garderai en mémoire la chaleur de leur accueil lors de nos réunions à Montpellier.

Je remercie également tous mes collègues de l'équipe IMEDIA. Un grand merci à Souheil Selmi pour son aide technique et son soutien amical. Je n'oublierai pas tous ces fous rires qui m'ont aidé à bien gérer la pression et le stress de la thèse. Merci également à Vera Bakic pour son rôle primordial dans le développement des outils d'évaluation et à Alexis Joly pour m'avoir autorisé à utiliser son algorithme de hachage. J'ai apprécié également les discussions avec Hervé Goëau et sa disponibilité. Enfin, j'ai vécu des moments privilégiés avec chacun des autres membres d'IMEDIA. Qu'ils trouvent ici l'expression de ma sincère amitié.

Je ne manquerai pas d'exprimer ma gratitude à tous mes enseignants universitaires et plus particulièrement à M. Ezzedine Zagrouba qui m'a initié au domaine du traitement d'images et de la vision par ordinateur. Il a accompagné mes premiers pas dans ce domaine et m'a encouragé à suivre le chemin de la recherche scientifique. Je le remercie pour son soutien et sa confiance.

Je ne trouve pas les mots pour remercier ma femme Yosra, source inépuisable d'énergie. Nous avons parcouru cette épreuve ensemble main dans la main. Yosra, tu as accompagné mes réussites avec beaucoup de joie et d'enthousiasme. J'ai aussi trouvé le réconfort auprès de toi dans les moments

difficiles. Si j'ai réussi à mener à bien cette thèse, c'est en grande partie grâce à toi. Je veux que tu trouves ici l'expression de mon plus grand amour et ma plus grande reconnaissance pour ton soutien durant cette thèse mais également pour avoir mis au monde notre fille Sarah.

Sarah, sans que tu le saches, tes regards et tes sourires innocents ont été une source de force et d'apaisement. Merci ma puce de m'avoir accompagné durant ces derniers mois.

Enfin, je ne remercierai jamais assez ma famille, ma mère et mes trois soeurs. Vous avez été toujours là pour m'épauler dans toutes les circonstances. Merci à mes deux nièces Amani et Hanin qui ont fait de mes vacances un vrai régal. Je vous aime tous et que Dieu vous garde pour continuer à remplir ma vie de bonheur.

Table des matières

Table des matières	iii
1 Introduction	1
2 Etat de l'art	9
2.1 Identification des plantes par les feuilles	10
2.2 Description de la forme d'une feuille	12
2.2.1 La forme de la feuille en botanique	12
2.2.2 Paramètres de forme	13
2.2.3 Description de la forme	18
2.3 Description de la nervation	30
2.3.1 Nervation en botanique	30
2.3.2 Extraction du réseau de nervures	30
2.3.3 Description du réseau de nervures	33
2.4 Description de la texture	35
2.4.1 Descripteurs <i>LBP</i>	35
2.4.2 Ondelettes	36
2.4.3 Autres descripteurs de texture	37
2.5 Description de la couleur des feuilles	37
2.5.1 Couleur des feuilles en botanique	37
2.5.2 Méthodes de description de couleur des feuilles	37
2.6 Comparaison et mise en correspondance	38
2.6.1 Recherche d'images par similarité (Image Retrieval)	39
2.6.2 Classification	42
2.7 Récapitulation et discussion	42
2.8 Choix et Hypothèses	44

2.8.1	Types d'images	44
2.8.2	Extraction du contour	45
2.8.3	Identification basée sur la recherche par similarité	46
3	Approches proposées	49
3.1	Description triangulaire de la forme	50
3.1.1	Représentation triangulaire entre des objets	51
3.1.2	Représentation triangulaire locale du contour	52
3.1.3	TAR (Triangle Area Representation)	56
3.1.4	Descripteurs de forme proposés	58
3.2	Contexte de forme étendu et points saillants	60
3.2.1	Points saillants d'une feuille en botanique	61
3.2.2	Détection des points saillants	62
3.2.3	Description des points saillants	66
3.3	Extraction et description de la nervation	71
3.3.1	Opérateurs morphologiques	72
3.3.2	Algorithme d'extraction de nervures	76
3.3.3	Résultats d'extraction des nervures	88
3.3.4	Description de la nervation	90
3.4	Conclusion	95
4	Evaluation	97
4.1	Évaluation sur la base Swedish	98
4.1.1	Caractéristiques de la base Swedish	98
4.1.2	Résultats des descripteurs de forme sur la base Swedish	99
4.1.3	Résultats du descripteur de contexte de forme étendu	105
4.1.4	Comparaison avec l'état de l'art	107
4.2	Évaluation sur la base Flavia	109
4.2.1	Caractéristiques de la base Flavia	109
4.2.2	Résultats des descripteurs de forme	111
4.2.3	Résultats du contexte de forme étendu (ASC)	116
4.2.4	Résultats des descripteurs de la nervation	119
4.2.5	Comparaison avec l'état de l'art	126
4.3	Évaluation sur la base One-hundred plant species leaves (Leaf100)	127

4.3.1	Caractéristiques de la base Leaf100	127
4.3.2	Résultats des descripteurs de forme	127
4.4	Évaluation sur les scans et les pseudo-scans de feuilles d'ImageCLEF 2011	131
4.4.1	Caractéristiques de la base ImageCLEF 2011 (Pl@ntLeaves)	131
4.4.2	Résultats des descripteurs de forme sur ImageCLEF2011	133
4.4.3	Résultats du contexte de forme étendu sur les scans et les pseudo-scans de test134	
4.4.4	Comparaison avec l'état de l'art	138
4.5	Evaluation sur les scans et les pseudo-scans de la base ImageCLEF 2012	140
4.5.1	Résultats des descripteurs de forme	141
4.5.2	Résultats du contexte de forme étendu	142
4.5.3	Comparaison avec l'état de l'art	143
4.6	Évaluation sur la base ImageCLEF2013	144
4.6.1	Résultats des descripteurs de forme	144
4.6.2	Résultats du contexte de forme étendu	145
4.6.3	Comparaison avec l'état de l'art	147
4.7	Conclusion	147
5	Fusion des descriptions foliaires	151
5.1	Méthodes de fusion	152
5.2	Fusion de la description de la forme et des points saillants	154
5.2.1	Comparaison et fusion des descripteurs sur la base Flavia	154
5.2.2	Fusion des descripteurs de nervation	157
5.3	Fusion de la description de la forme et de la nervation	159
5.3.1	Comparaison entre les descriptions de forme et de nervures	160
5.3.2	Fusion des descripteurs de forme et de nervation	161
5.4	Requêtes multiples	162
5.4.1	Évaluation sur ImageCLEF 2012	162
5.4.2	Évaluation sur ImageCLEF 2013	166
5.5	Comparaison avec l'état de l'art	168
5.5.1	Comparaison sur la base ImageCLEF 2012	168
5.5.2	Comparaison sur la base ImageCLEF 2013	168
5.5.3	Comparaison sur la base Flavia	169
5.6	Conclusion	171

<i>TABLE DES MATIÈRES</i>	vi
6 Conclusion générale et perspectives	173
A Résultats d'extraction de nervures sur la base Flavia	179
Bibliographie	189

Résumé

IL est nécessaire de reconnaître les espèces végétales afin de préserver la biodiversité des écosystèmes. L'identification d'une plante, habituellement effectuée par les experts, se base sur l'observation de ses organes et en majeure partie sur des critères visuels. La feuille est l'organe le plus utilisé grâce à sa facilité d'acquisition. De plus, celle-ci contient une information importante sur la taxonomie de la plante. Ceci permet d'envisager d'utiliser l'analyse d'images pour élaborer un processus de reconnaissance automatique de l'espèce végétale à partir de la donnée d'une image de feuille.

Nous introduisons une nouvelle approche d'identification d'espèces végétales, basée sur la description des caractères foliaires suivants : la forme, les points saillants et la nervation.

En premier lieu, la forme de la feuille est représentée par des descripteurs locaux associés aux points échantillonnés sur le contour. Différentes représentations triangulaires multi-échelle sont introduites et comparées. Pour décrire les points saillants de la feuille, nous proposons une représentation dérivée du contexte de forme (Shape Context). Finalement, la nervation est extraite par la détection de structures linéaires élémentaires avec des outils morphologiques. Le réseau de nervures extrait est décrit par ses directions principales et sa répartition spatiale dans le contexte de la surface de la feuille.

Pour tous les descripteurs, une méthode de mise en correspondance locale est utilisée. Des évaluations, menées sur six bases de feuilles publiques, montrent que nos approches permettent généralement d'identifier l'espèce végétale de la feuille et que la fusion tardive des descripteurs augmente la précision de l'identification.

Mots-clés : identification des espèces végétales, descripteur de forme 2D, représentation triangulaire multi-échelle, contexte de forme, mise en correspondance locale, fusion tardive.

Abstract

Plant species identification, usually performed by specialists, is based on the observation of its organs and mostly on visual criteria. Thanks to its ease of acquisition, the leaf is the most used organ. In addition, it contains important information on the taxonomy of the plant. This enables the use of computer vision and pattern recognition techniques for developing an automatic recognition process of the plant species from a leaf image. We introduce a new approach to identify plant species, based on the description of the following leaf characteristics : its shape, its saillant points and its venation. First, the shape of the leaf is represented by local descriptors associated to a set of points sampled on the contour. Different multi-scale triangular representations are introduced and compared. To describe the saillant points of the leaf, we propose a shape context based representation. Finally, the venation is extracted by detecting elementary linear structures with morphological tools. The venation network is described by its main directions and its spatial distribution in the context of the leaf boundary. A local matching method is used for all descriptors. Evaluations, conducted on six publicly available plant identification benchmarks, show that our approaches identify the plant species of the leaf in most of the cases and that the late fusion of the proposed descriptors improves the identification process.

Key words : plant species identification, 2D shape descriptor, multi-scale triangular representation, shape context, local features matching, late fusion.

Chapitre 1

Introduction

L'intérêt pour l'économie de l'environnement a beaucoup augmenté ces dernières années. Cela s'est traduit, en particulier, par un nombre important de travaux dédiés à l'étude ou à la classification des espèces vivantes. Différents outils numériques pour la classification des espèces animales à partir d'une base d'images ont été proposés pour les oiseaux [BLL⁺14, HMLZ14], les chiens [LKJB12], les papillons [KK14], les insectes [BPC⁺12], les poissons [SGDS⁺10], etc. D'autres travaux se sont intéressés à l'étude des espèces végétales à partir des images de plantes.

Les plantes jouent un rôle primordial dans l'équilibre des écosystèmes et d'une manière plus générale, dans la préservation de toute forme de vie sur notre planète. En effet, les plantes sont à l'origine de l'oxygène contenu dans l'atmosphère. Elles constituent la nourriture d'un très grand nombre d'espèces vivantes et représentent ainsi un élément fondamental de la chaîne trophique (chaîne alimentaire). Selon une récente étude publiée dans la revue PLoS Biology, Mora et al. [MTA⁺11] confirment que la terre abriterait environ 298 000 espèces de plantes dont 215 644 ont été décrites et cataloguées. Cela signifie que près de 30% des espèces végétales connues restent non classées sans compter le grand nombre d'espèces qui ne sont pas encore découvertes et qui ont besoin d'être identifiées.

Identifier une plante consiste à déterminer son espèce végétale. Ce processus est traditionnellement effectué par un botaniste ou un naturaliste. Ceux-ci se basent sur l'observation de la plante et la connaissance des clefs de détermination. Une liste de caractères botaniques est présentée dans [Ell09]. La majorité de ces caractéristiques peuvent être associées à des attributs visuels comme c'est le cas de la feuille Mimosa de la figure 1.1.

Avec l'utilisation généralisée des appareils photographiques et l'explosion du nombre d'images sur internet, les botanistes ne peuvent répondre à toutes les demandes d'identification. L'existence d'outils numériques permettant d'assister ou de remplacer le botaniste et de traiter un grand nombre de données s'avère très utile. Certains botanistes du CIRAD ont proposé des outils interactifs descriptifs, comme l'outil IDAO (IDAO identification assisted by computer)¹ (cf. Figure 1.1). Les méthodes numériques d'identification d'espèces se basent sur l'analyse du contenu visuel des images. Elles permettent d'automatiser ou d'améliorer le processus d'identification.

1. <http://idao.cirad.fr>

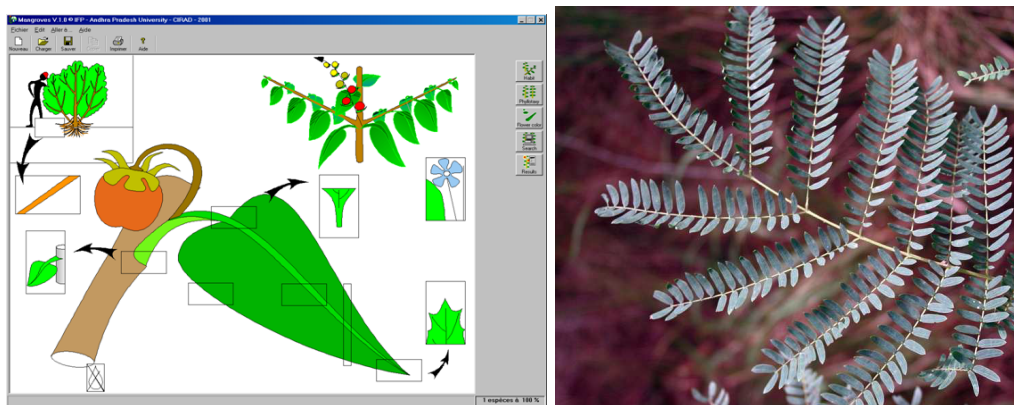


FIGURE 1.1 – Image de feuille de l'espèce *Mimosa glauca* L. Feuille alterne, composée, bipennée de 15 à 25 cm de long. Présence de 3 à 7 paires de pennes opposées de 4 à 8 cm de long, composées d'une douzaine de paires de folioles linéaires-oblongues, ciliolées à sommet aigu, de 0.7 à 2 cm de long et 2 à 5 mm de large. Images et description tirées de l'outil IDAO.

Différents projets de recherche proposent des applications d'identification d'images de plantes, mobiles ou disponibles sur internet :

- Le système embarqué d'identification d'espèces végétales CLOVER [NHK05].
- Le projet américain LeafSnap, au sein duquel une application mobile d'identification de plantes [KBB⁺12] a été développée via une collaboration entre l'université de Columbia et l'université de Maryland et dont la première version a été présentée dans [BCF⁺08].
- Le projet Pl@ntNet² qui vise à mettre en œuvre une plateforme de partage de connaissances en botanique.
- Le projet ReVes³[CTM⁺13, Cer13] qui a abouti au développement de l'application mobile Folia dédiée à la reconnaissance des espèces végétales.
- L'application mobile Flora Finder⁴ développée par MEA Mobile qui vise à identifier des espèces végétales natives en Nouvelle Zélande [WPDP12].
- La plateforme IHerbarium⁵ qui propose une application iPhone de collecte d'images et d'aide à l'identification.

Cette liste n'est pas exhaustive mais montre le grand intérêt du public et des professionnels autour de cette problématique.

2. projet financé par Agropolis Fondation, voir <http://www.plantnet-project.org>

3. <http://liris.univ-lyon2.fr/reves/>

4. <http://www.otago.ac.nz/botany/florafinder/>

5. <http://www.iherbarium.fr/>

Projet Pl@ntNet

Notre travail de thèse a été effectué dans le cadre du projet Pl@ntNet d'Agropolis Fondation. Ce projet est une initiative proposée par un consortium international dont les acteurs principaux sont :

- **AMAP** : une unité mixte de recherche (botAnique et bioinforMatique de l'Architecture des Plantes"). Elle regroupe des chercheurs et des botanistes de plusieurs organismes (Cirad, Cnrs, Inra, Ird et l'université Montpellier 2). Ses travaux sont dédiés à des problématiques liées à l'agronomie, la botanique et la foresterie.
- **L'équipe-projet IMEDIA** de l'Inria Paris-Rocquencourt dont les activités principales sont le développement de nouvelles approches pour d'indexation, la structuration efficace du contenu visuel, la recherche interactive et la navigation dans des bases d'images, de vidéos ou de modèles 3D.
- **Le réseau associatif Tela-Botanica**⁶ qui est en collaboration directe avec des chercheurs, des professionnels et le grand public. Il développe des outils collaboratifs afin de faciliter le partage d'informations et de connaissances entre les botanistes francophones.

Au sein du projet Pl@ntNet, des collectes de données ont été effectuées, lors d'études botaniques spécifiques ou via l'outil "Carnet en ligne" proposé par le réseau Tela-Botanica. Elles ont permis de créer et d'enrichir des bases d'images botaniques annotées utilisées par l'application Pl@ntNet-Identify⁷ développée à partir du moteur de recherche Ikona-maestro [BFF⁺01].

Notre objectif ici est d'introduire une approche innovante d'identification d'espèces par l'analyse du contenu visuel d'images de plantes. Nous avons choisi d'explorer des méthodes basées sur des modèles géométriques, intégrant notamment des descriptions locales dans les images. Ces méthodes doivent :

- déterminer l'espèce végétale d'une feuille donnée en entrée i.e la bonne espèce doit figurer dans les premières positions dans la liste des espèces retournée en réponse,
- être génériques et robustes aux variations de forme des feuilles d'une même espèce végétale,

6. <http://www.tela-botanica.org/>

7. <http://identify.plantnet-project.org/fr/>

- être intégrées dans un moteur de recherche d’images par le contenu [BFF⁺01].

Pour répondre à ces exigences, nous avons fait plusieurs propositions et études expérimentales que nous pouvons présenter sous forme de contributions.

Description triangulaire de la forme de la feuille

Cette première contribution porte sur la description de la forme foliaire. Une famille de descripteurs est introduite. Ceux-ci permettent d’effectuer une représentation locale ou globale du contour de la feuille en fonctions des paramètres utilisés. Nous avons utilisé une méthode de représentation du contour par des triangles à plusieurs échelles.

Corrélation spatiale entre les points saillants et le contour de la feuille

Une deuxième famille de descripteurs vise à représenter les points saillants de la feuille via leurs relations spatiales ou vis-à-vis du contour. Pour ce faire, une extension du descripteur contexte de forme (Shape Context) [BMP02] est proposée.

Extraction et description de la nervation

Un troisième ensemble de descripteurs est dédié à la représentation de la nervation de la feuille. Ceci nécessite d’extraire le plus finement possible le réseau de nervures d’une image de feuille puis le décrire.

Évaluation expérimentale approfondie

Nous avons effectué une évaluation systématique de nos descripteurs sur les principales bases d’images de feuilles publiques. Pour chaque base, nous avons adopté les mesures de performance utilisées par les approches existantes pour pouvoir comparer nos résultats.

Ce mémoire de thèse est structuré de la façon suivante :

- En premier lieu, nous passons en revue les différentes techniques introduites dans la littérature pour décrire une feuille dans le chapitre 2. Les méthodes sont classées selon le caractère foliaire utilisé.
- Dans le chapitre suivant, nous explicitons l’ensemble des approches proposées.

- Le chapitre 4 est consacré à l'évaluation des descripteurs foliaires. Six base d'images de feuilles publiques sont utilisées : Swedish [Söd01], Flavia [WBX⁺07], One-hundred plant species (Leaf 100)[MCO13] et les bases ImageCLEF en 2011 [GBJ⁺11], 2012 [GBJ⁺12] et 2013 [GJB⁺13].
- Une description plus complète d'une feuille, résultant de la combinaison des descriptions précédentes par fusion, est proposée et évaluée dans le cinquième chapitre.
- Finalement, en conclusion, nous dressons un bilan de nos contribution et proposons des extensions de ce travail de thèse.
- Une annexe complète ce mémoire. Elle présente des résultats d'extraction de nervures sur les images de feuilles de la base Flavia [WBX⁺07].

LISTE DES PUBLICATIONS

Raffi Enficiaud et Sofiène Mouine. **Landmark Extraction from Leaves with Palmate Venation - Application to Grape**. International Conference on Pattern Recognition - Applications and Methods (2012).

Sofiène Mouine, Itheri Yahiaoui et Anne Verroust-Blondet. **Advanced shape context for plant species identification using leaf image retrieval**. ACM International Conference on Multimedia Retrieval (2012).

Vera Bakic, Itheri Yahiaoui, Sofiène Mouine, Saloua Litayem, Wajih Ouertani, Anne Verroust-Blondet, Hervé Goëau, Alexis Joly. **Inria IMEDIA2's Participation at ImageCLEF 2012 Plant Identification Task**. CLEF (Online Working Notes/Labs/Workshop) 2012.

Sofiène Mouine, Itheri Yahiaoui, Anne Verroust-Blondet. **A shape-based approach for leaf classification using multiscaletriangular representation**. ACM International Conference on Multimedia Retrieval (2013).

Sofiène Mouine, Itheri Yahiaoui, Anne Verroust-Blondet, Laurent Joyeux, Souheil Selmi, Hervé Goëau. **An android application for leaf-based plant identification**. ICMR 2013.

Sofiène Mouine, Itheri Yahiaoui, Anne Verroust-Blondet. **Plant species recognition using spatial correlation between the leaf margin and the leaf salient points**. International Conference on Image Processing (2013).

Sofiène Mouine, Itheri Yahiaoui, Anne Verroust-Blondet. **Combining Leaf Salient Points and Leaf Contour Descriptions for Plant Species Recognition**. International Conference on Image Analysis and Recognition (2013).

Vera Bakic, Sofiène Mouine, Saloua Ouertani-Litayem, Anne Verroust-Blondet, Itheri Yahiaoui, Hervé Goëau, Alexis Joly. **Inria's Participation at ImageCLEF 2013 Plant Identification Task**. CLEF (Working Notes) (2013).

Chapitre 2

Etat de l'art

Le processus d'identification de plantes s'effectue le plus souvent par l'observation d'un ou plusieurs organes (feuille, fleur, écorce, *etc.*). Chaque organe possède des caractères particuliers lui permettant d'être une source d'information déterminante pour trouver l'espèce correspondante à la plante. Le choix de l'organe de base dans le processus d'identification est guidé par certains critères :

- la quantité d'information taxonomique pertinente offerte par l'organe. Ce critère est purement botanique et permet de répondre à la question suivante : Quel est pour un expert l'organe le plus déterminant pour identifier une plante ?
- la présence de l'organe sur la plante, par exemple, une feuille est présente plus longtemps qu'un fruit.
- il faut que le protocole d'acquisition soit simple à mettre en œuvre par l'utilisation des appareils d'acquisition appropriés.

Nous nous intéressons aux approches numériques permettant d'effectuer l'identification en se basant sur l'information visuelle disponible.

La plupart des méthodes publiées se basent sur l'analyse du contenu visuel des images de feuilles pour la reconnaissance de plantes. Une feuille a l'avantage d'être accessible et facilement acquise utilisant un appareil de photographie ou un scanner. De plus, elle possède des caractères qui représentent souvent des indices clés pour l'identification : la forme, la nervation ou la texture.

D'autres approches proposées sont basées sur l'étude d'autres organes qui ne sont pas toujours présents sur la plante ou qui ne sont pas forcément déterminants : les fleurs [DMR99, HCL⁺04, NZ08, Nil09]; les racines [HJSS92]; les écorces [HHD⁺06]; les fruits [ASS⁺11]; plusieurs organes à la fois [GBB⁺12] (cf. [CCC⁺12] pour un état de l'art sur le domaine).

Nous avons choisi de considérer le contenu foliaire comme base de notre système d'identification. Nous consacrons ce chapitre à la présentation des méthodes d'identification de plantes par les feuilles. Nous y détaillerons les différentes orientations des chercheurs pour la caractérisation des feuilles dans un système d'identification automatique de plantes.

2.1 IDENTIFICATION DES PLANTES PAR LES FEUILLES

Les méthodes d'identification de plantes basées sur les feuilles peuvent être classées en fonction du caractère foliaire utilisé (cf. Figure 2.1) :

- la forme : cette catégorie comporte deux sous catégories, des descripteurs génériques de forme 2D qui ont été testés sur les images de feuilles et des mesures géométriques spécifiques à la feuille, basées sur des connaissances botaniques.
- La nervation : elle constitue un caractère important de la feuille. Les approches existantes extraient en premier lieu le réseau de nervures puis proposent de le décrire et d'étudier son impact dans l'identification de l'espèce.
- la texture : les approches basées sur la texture de la feuille représentent les structures locales du limbe. Les méthodes les plus courantes utilisent les ondelettes, les filtres de Gabor et la dimension fractale. Un deuxième groupe d'approches propose de nombreuses variantes du descripteur LBP (Local Binary Patterns).
- la couleur : il s'agit d'extraire des attributs colorimétriques à partir des images de feuilles en couleur.

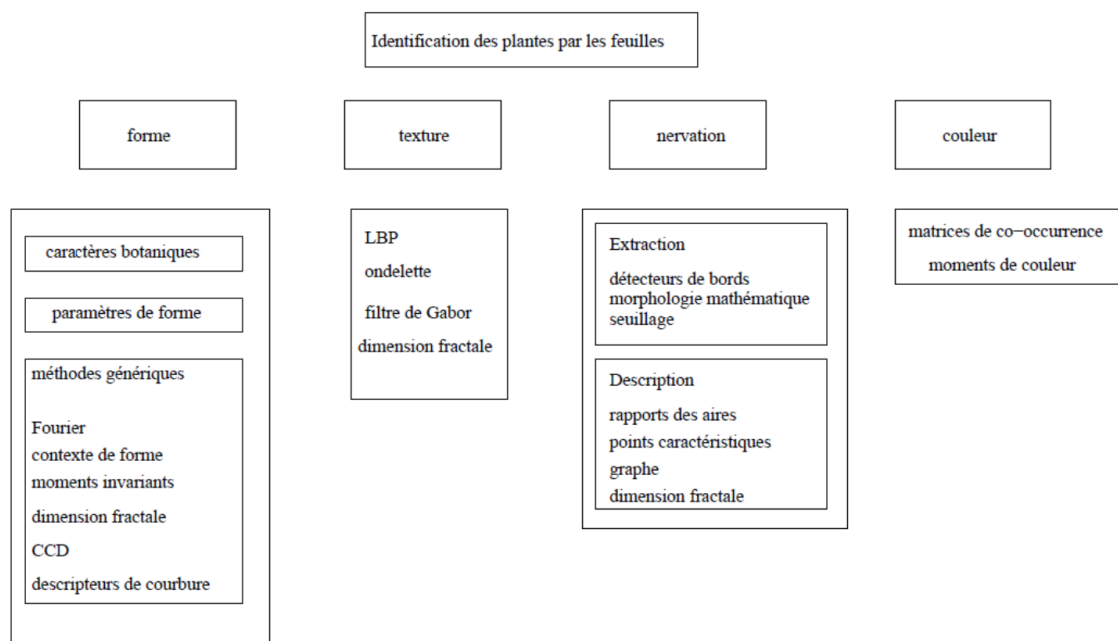


FIGURE 2.1 – Classification des méthodes d'identification des plantes par les feuilles

Dans les sections suivantes, nous nous focalisons sur la façon dont chaque caractère est décrit dans la littérature.

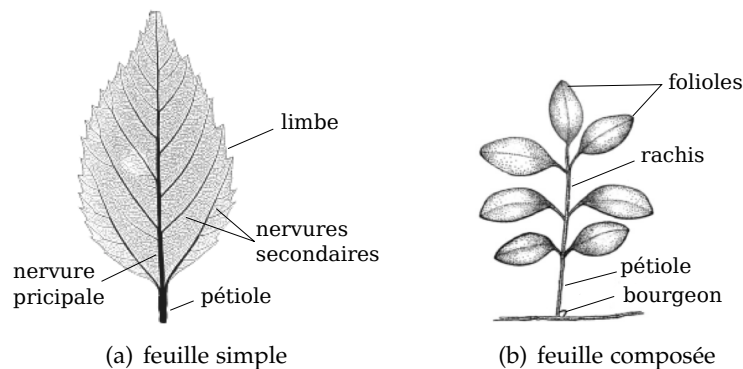


FIGURE 2.2 – Feuille simple et feuille composée

2.2 DESCRIPTION DE LA FORME D'UNE FEUILLE

2.2.1 La forme de la feuille en botanique

La forme de la feuille est une des caractéristiques les plus visibles et les plus stables. Cela explique sa large utilisation dans le contexte d'identification des espèces végétales. Deux grands types de feuilles sont distingués selon la forme globale (cf. Figure 2.2) :

- les feuilles simples qui possèdent un seul limbe continu. Un bourgeon est présent à l'extrémité du pétiole. Des exemples de formes de feuilles simples sont montrés dans la figure 2.4
- les feuilles composées qui possèdent plusieurs folioles se caractérisant par l'absence de bourgeon à leur base, le bourgeon se trouve à la base du pétiole.

Ceci peut représenter une difficulté à prendre en compte pour les descripteurs de forme foliaire car un bon descripteur de feuilles simples n'a pas forcément la même performance sur les feuilles composées et vice versa. Un autre défi majeur est la similarité inter-classe et la variabilité intra-classe des formes de feuilles (cf. Figure 2.3). Ce problème est un défi y compris pour l'homme. Une expérimentation intéressante a été effectuée dans ce contexte. Elle a consisté à proposer vingt images de vingt variétés de vigne différentes, très similaires au niveau de la forme, à un expert pour les identifier. Cette expérimentation a été répétée trois fois en faisant varier la nature des images :

- des images originales
- des images du contour de la feuille
- des images du contour + nervures

Le résultat est reporté dans le tableau 2.1.

Nous pouvons conclure que la difficulté que représente la similarité inter-classe est non



FIGURE 2.3 – Similarité inter-classe et variabilité intra-classe. Première ligne : similarité inter-classe des variétés de vigne : Melon, Pinot et Gamay. Deuxième ligne : variabilité intra-classe de la variété *Ficus carica*

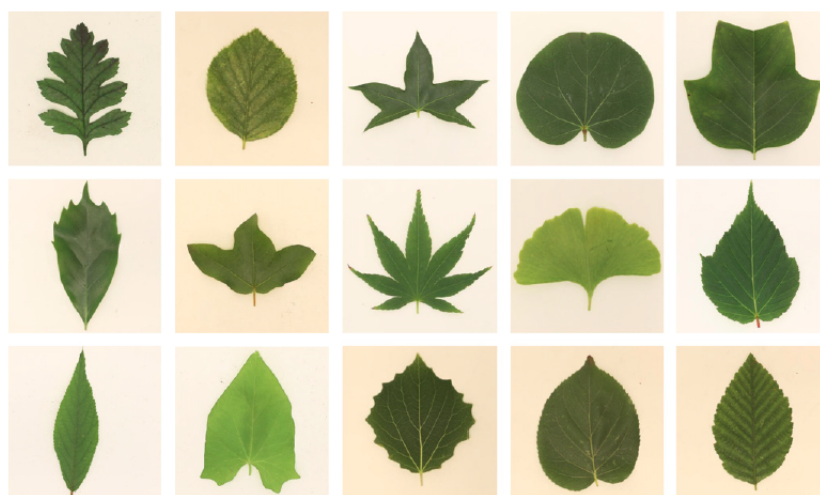


FIGURE 2.4 – exemples de formes foliaires

négligeable, même pour un expert. Généralement, une information supplémentaire sur d'autres caractères de la feuille est nécessaire.

	Images originales	Contour	Contour + nervures
Pourcentage d'identification	93%	60%	86%

TABLE 2.1 – Résultats d'identification des variétés de vigne par un expert

2.2.2 Paramètres de forme

Un paramètre de forme est une mesure géométrique simple et globale de la forme. Nous distinguons les paramètres de forme **génériques** [MKJ08] qui peuvent s'appliquer à n'importe quelle forme et les paramètres de forme **spécifiques** qui sont issus des connaissances

en botanique et qui calculent des mesures spécifiques aux feuilles.

Le paragraphe suivant est consacré à la définition des concepts de base nécessaires pour introduire les paramètres de forme.

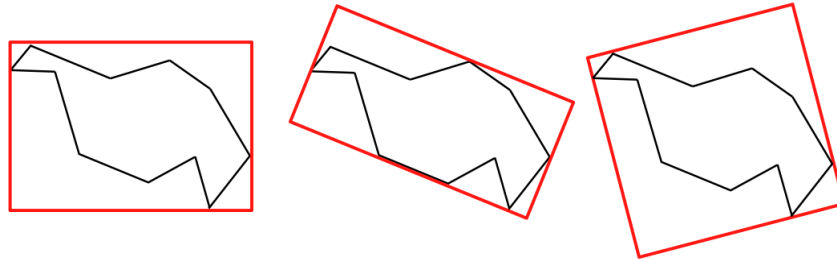


FIGURE 2.5 – Différents types de boîtes englobantes. De gauche à droite : boîte englobante, boîte englobante minimale et boîte englobante maximale

Concepts de base

Forme

Une forme désigne une région R du plan, généralement connexe, définie dans un repère donné $\mathcal{R}(\mathcal{X}, \mathcal{Y})$. R divise le plan en deux parties : une partie interne à la région R et une partie externe à R qui correspond au fond. Dans la suite, nous utiliserons d'une façon équivalente les termes forme et région pour désigner R .

Contour d'une forme

Le contour d'une forme $Cont(R)$ est défini par l'ensemble des points qui représentent la frontière séparant R du fond.

Boîte englobante

Elle est définie par un rectangle aligné $BE(R)$ contenant la région R , défini par ses dimensions $x_{max} - x_{min}$ et $y_{max} - y_{min}$ (cf. Figure 2.5) avec :

x_{max} : l'abscisse maximale de tous les points du contour de la forme.

x_{min} : l'abscisse minimale de tous les points du contour de la forme.

y_{max} : l'ordonnée maximale de tous les points du contour de la forme.

y_{min} : l'ordonnée minimale de tous les points du contour de la forme.

Notons que la boîte englobante n'est pas invariante à la rotation de R . De même, la largeur

l et la hauteur h d'une forme sont définies de la manière suivante :

$$l = x_{max} - x_{min} \quad \text{et} \quad h = y_{max} - y_{min} \quad (2.1)$$

Boîte englobante minimale

Elle est définie par un rectangle orienté $BE_{Min}(R)$ d'aire minimale contenant R (cf. Figure 2.5).

Boîte englobante maximale

Une boîte englobante maximale est définie par le rectangle orienté $BE_{Max}(R)$ d'aire maximale contenant R et dont chaque côté touche le contour externe en au moins un point (cf. Figure 2.5).

Enveloppe convexe

L'enveloppe convexe $EC(R)$ correspond au plus petit ensemble convexe qui contient R (cf. Figure 2.6(a)). Si R est convexe, R est alors égale à son enveloppe convexe.

Cercle inscrit

Le cercle inscrit de R est le cercle maximal $C_{insc}(R)$ de rayon r_i contenu dans R (cf. Figure 2.6(b)).

Cercle circonscrit

Le cercle circonscrit de R est le cercle minimal $C_{circ}(R)$ de rayon r_c contenant R (cf. Figure 2.6(c)).

Paramètres de forme génériques

Les paramètres de forme les plus courants dans la littérature sont les suivants :

- Aire (Area) [ASFRRGG11, CA10, WBX⁺07] : généralement exprimée en nombre de pixels appartenant à la région.

$$\mathcal{A}(R) = \text{card}(\mathcal{P}) \quad \text{où} \quad \mathcal{P} = \{p(x, y) \in R\} \quad (2.2)$$

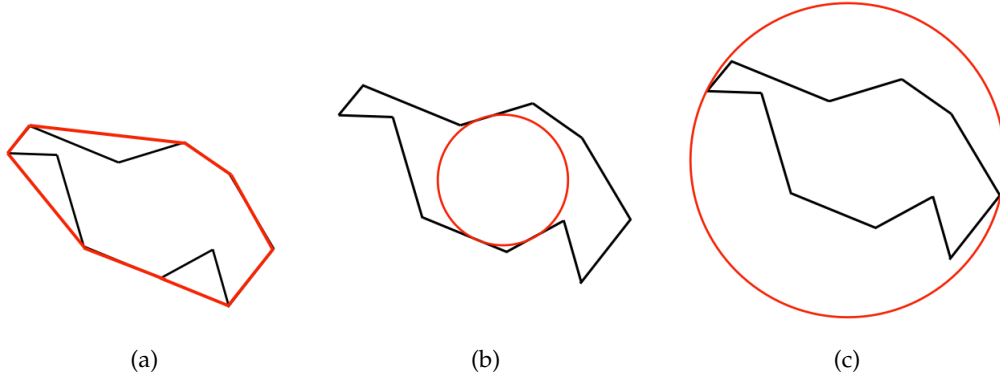


FIGURE 2.6 – Enveloppe convexe et cercles (inscrit et circonscrit) d'une forme. (a) enveloppe convexe (b) cercle inscrit (c) cercle circonscrit

- Périmètre (Perimeter) [ASFRRGG11, WBX⁺07] : nombre de points du contour de la région $Cont(R)$.

$$Perimeter(R) = card(Cont(R)) \quad (2.3)$$

- Diamètre (Diameter) [WBX⁺07, CA10] : correspond à la distance maximale séparant une paire de points distincts (p_i, p_j) du contour $Cont(R)$.

$$Diameter(R) = \max_{p_i \in Cont(R), p_j \in Cont(R)} ||p_j - p_i|| \quad (2.4)$$

- Rectangularité (Rectangularity) [WBX⁺07, YMB12] : c'est le rapport des aires de la région R et de sa boîte englobante $BE(R)$.

$$Rectangularity(R) = \frac{\mathcal{A}(R)}{\mathcal{A}[BE(R)]} \quad (2.5)$$

- Sphéricité (Sphericity) [KNSS11a, YMB12] : correspond au rapport entre le rayon r_c du cercle circonscrit $C_{circ}(R)$ et le rayon r_i du cercle inscrit $C_{insc}(R)$ de la région R .

$$Circularity(R) = \frac{r_i(R)}{r_c(R)} \quad (2.6)$$

- Excentricité (Eccentricity) [ERAK⁺05, GPRG12, KNSS11a, WBX⁺07] : définie par le rapport entre la longueur de l'axe d'inertie et celle de l'axe secondaire. Ceci équivaut au calcul du rapport entre la largeur l et la hauteur h de la boîte englobante minimale de la région $BE_{Min}(R)$. Une autre mesure, appelée Élongation, basée sur l'excentricité, est définie par l'équation 2.8.

$$Eccentricity(R) = \frac{l}{h} \quad (2.7)$$

$$Elongation(R) = 1 - Eccentricity(R) \quad (2.8)$$

- Circularité (Circularity) [ERAK⁺05, MA04, SK00, WBX⁺07, YMB12] : est le rapport des aires de la région R et du cercle C ayant le même périmètre que R .

$$Circularity(R) = \frac{\mathcal{A}(R)}{\mathcal{A}(C)} \text{ avec } Perimeter(R) = Perimeter(C) \quad (2.9)$$

- Solidité (Solidity) [ERAK⁺05, YMB12] : mesure dénommée aussi Convex Area Ratio. Elle correspond au rapport entre l'aire $\mathcal{A}(\mathcal{R})$ de la région et celle de son enveloppe convexe $\mathcal{A}[EC(R)]$.

$$CAR(R) = \frac{\mathcal{A}(R)}{\mathcal{A}[EC(R)]} \quad (2.10)$$

- Rapport des périmètres (Convex Perimeter Ratio) : correspond au rapport entre le périmètre de la région $Perimeter(R)$ et de celui de son enveloppe convexe $Perimeter[EC(R)]$.

$$CPR(R) = \frac{Perimeter(R)}{Perimeter[EC(R)]} \quad (2.11)$$

- Aspect Ratio [DWZ07, ERAK⁺05, GPRG12, SK00, WBX⁺07] : rapport entre la boîte englobante minimale $BE_{Min}(R)$ de dimension $l \times h$ et la boîte englobante maximale $BE_{Max}(R)$ de dimension $L \times H$

$$AR(R) = \frac{L \times H}{l \times h} \quad (2.12)$$

Paramètres de forme spécifiques

Des paramètres de forme spécifiques, significatifs pour les botanistes ont été introduits. Deux points caractéristiques sont souvent utilisés dans ce type de paramètres :

- L'apex : le sommet de la feuille, il est généralement identifié par la terminaison de la nervure principale de la feuille.
- la base : le point d'attache entre le pétiole et le limbe.

Différentes techniques ont été proposées pour la détection automatique de ces deux points particuliers [MYB12, RSBG13]. La figure 2.7 montre quelques distances spécifiques à la

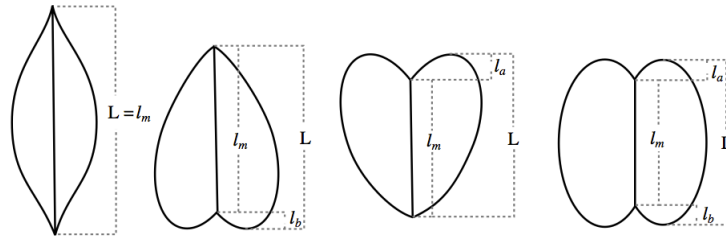


FIGURE 2.7 – Distances spécifiques à la feuille. L désigne la longueur du limbe, l_m est la longueur de la nervure centrale principale, l_a est la distance séparant l'apex et le point supérieur maximal de la feuille et l_b est la distance entre la base et le point inférieur maximal de la feuille. Notons que $L = l_a + l_m + l_b$

feuille. D'une façon similaire, les paramètres spécifiques peuvent décrire les angles à l'apex, la forme basale, la dentition et les lobes, *etc.*

2.2.3 Description de la forme

Descripteurs morphologiques

Un descripteur morphologique est un vecteur décrivant la morphologie d'une forme à travers le codage d'un ou plusieurs paramètres de forme. Ceci peut être utilisé comme :

1- Une étape de filtrage [DWZ07, WCFW00, WCF03, CA10] :

- pour réduire l'espace de recherche dans un premier temps, c'est-à-dire exclure *a priori* les images de feuilles qui sont très différentes.
- en combinaison avec d'autres types de descripteurs foliaires afin d'apporter plus de précision lors de la caractérisation de la feuille.

C'est le cas le plus fréquent dans la littérature.

2- Base de l'identification. Dans ce cas, il s'agit d'une concaténation de paramètres de forme (génériques ou spécifiques) ou un mélange entre ces deux types.

Examinons les travaux de la littérature qui utilisent les paramètres de forme pour procéder à un premier filtrage des images afin d'éliminer les images fortement différentes par rapport à une image requête.

- L'excentricité permet à Wang et al. [WCFW00, WCF03] de réduire l'espace de recherche dans une première étape. Elle est ensuite retenue parmi d'autres descripteurs pour calculer un score de classification dans la deuxième étape de l'algorithme. L'approche globale a été évaluée sur des feuilles de plantes médicinales. Cette base privée comporte 140 espèces

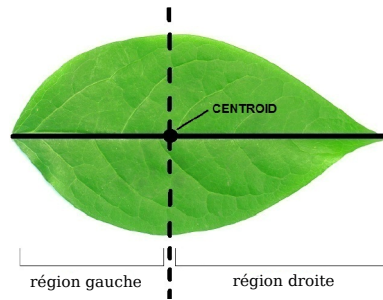


FIGURE 2.8 – Deux régions considérées dans [UU12]. Figure extraite de [UU12]

contenant chacune dix images.

- Une démarche similaire a été proposée par Caballero et Aranda [CA10]. Une première étape de l'algorithme consiste à réduire l'espace de recherche en utilisant trois descripteurs morphologiques, à savoir : le rapport des aires de la région et de sa boîte englobante minimale, l'excentricité et le diamètre. Chacun de ces descripteurs est testé avec différentes valeurs de seuil comprises entre 0 et 1. Le meilleur seuil pour chaque descripteur est retenu en fonction de trois mesures de pertinence des résultats : précision, rappel et pourcentage de réduction de la base. Un descripteur de courbure est ensuite calculé sur l'ensemble d'images résultant de la première étape. Pour évaluer cette méthode, les auteurs utilisent une base de feuilles privée qui compte 300 images et 13 classes.

Du et al. [DWZ07] calculent neuf descripteurs morphologiques basés sur les paramètres de forme. En plus de la description morphologique, les auteurs calculent des moments invariants (voir paragraphe 2.2.3). Cette approche a été évaluée sur une base d'images de feuilles privée comportant 400 images réparties en 20 classes.

Wang et Gao [WG13] combinent des descripteurs globaux de forme avec un descripteur local de courbure. Pour décrire globalement la forme de la feuille, trois descripteurs morphologiques sont utilisés : solidité, excentricité et rectangularité. Un score de dissimilarité est défini à partir de ces trois descripteurs pour augmenter la précision de la classification. Les expérimentations sont conduites sur la base Swedish [Söd01] et sur une base privée qui comporte 1200 images de feuilles appartenant à 100 espèces. Sur la base Swedish, un protocole d'évaluation (test Bullseye), rarement calculé pour cette base, a été utilisé. Les résultats sont très proches de ceux de l'état de l'art mais le point fort de cette méthode est surtout le temps de calcul réduit. D'ailleurs, ceci représente un avantage majeur de l'utilisation des descripteurs morphologiques.

Uluturk et Ugur [UU12] utilisent exclusivement les paramètres de forme pour décrire la forme foliaire. Ils comparent deux sous régions de la feuille au lieu de travailler sur son

intégralité (cf. Figure 2.8). Trois descripteurs morphologiques sont calculés sur chaque sous région, à savoir, le rapport des valeurs des aires, des rectangularités et des proportions. Ces trois descripteurs sont enrichis par sept descripteurs morphologiques classiques calculés sur toute la feuille. L'évaluation de cette approche est effectuée sur la base Flavia [WBX⁺07]. Un score d'identification de 92% a été obtenu. Tenant compte des caractéristiques de la base Flavia, le score de 92% est un bon score mais sachant que les auteurs utilisent un ensemble d'images d'apprentissage sept fois plus grand que l'ensemble d'images de test pour chaque espèce, le score affiché ne reflète pas forcément la performance de cette approche.

Les descripteurs morphologiques sont extrêmement efficaces pour repérer les formes qui sont très différentes d'une forme requête. Cela explique l'utilisation fréquente de ces descripteurs pour réduire l'espace de recherche malgré le risque de propagation d'erreur. Comme la majorité de ces descripteurs se basent sur le calcul de rapports de distances et d'aires, ils sont invariants à la translation, à la rotation et, quand l'espace est normalisé, à l'échelle. De plus, ils sont très faciles à calculer ce qui leur donne une importance accrue, en particulier pour l'identification ou la classification des formes foliaires.

Cependant, les résultats obtenus par ces descripteurs dans la littérature montrent qu'ils sont loin d'atteindre une précision très élevée. En effet, leur robustesse aux déformations non-rigides et à l'occultation partielle est faible. De plus, ils sont très sensibles à la similarité inter-classe et la variabilité intra-classe. C'est pourquoi les descripteurs morphologiques sont généralement combinés avec d'autres types de descripteurs.

Descripteurs géométriques

Descripteurs de courbure

Les descripteurs de courbure représentent la forme en terme de concavités/convexités de son contour. D'une façon générale, si nous considérons un repère orthonormé et une courbe plane paramétrée $\gamma(t)(x(t), y(t))$, la courbure est définie par :

$$\gamma(t) = \frac{x'(t).y''(t) - x''(t).y'(t)}{(x'^2(t) + y'^2(t))^{\frac{3}{2}}} \quad (2.13)$$

La méthode de représentation de courbure CSS (Curvature Scale Space), proposée par Mokhtarian et al. [MAK96], consiste à calculer les valeurs de courbure sur des contours à différentes échelles et à différents niveaux de lissage. Le lissage est proportionnel à l'écart

type σ des gaussiennes successives appliquées au contour comme illustré dans la figure 2.10. Le descripteur CSS permet donc de décrire les convexités (par une courbure positive) et les concavités (par une courbure négative) d'une forme.

Un CSS amélioré a été présenté dans [MA04] pour la classification des feuilles. Dans cette méthode, les auteurs abordent le problème d'auto-intersection du contour qui se retrouve souvent dans le cas des feuilles lobées ou composées à cause du recouvrement des lobes ou des folioles.

Cette approche permet, de plus, de représenter le contour par un ensemble de points caractéristiques correspondant aux extrema de courbure [CA10] (cf. Figure 2.9). Cerruti et al. [CTCV13] introduisent une description structurale du contour dérivée du CSS. Ils calculent la courbure d'une façon différente et sans lissage gaussien.

Une autre description de courbure est introduite par Kumar et al. [KBB⁺12] dans leur

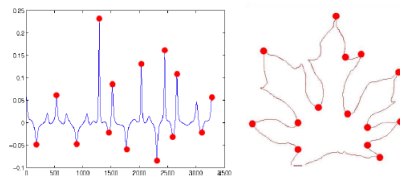


FIGURE 2.9 – Détection des points caractéristiques du contour dans [CA10] par l'application de l'algorithme CSS. Points extraites à une échelle donnée $s = 11$. Figure extraite de [CA10]

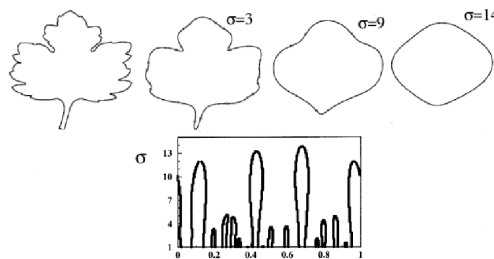


FIGURE 2.10 – Description de courbure dans l'espace courbure-échelle (Curvature Scale Space). Figure extraite de [MA04]

système d'identification de plantes sur mobile *LeafSnap*. Ils considèrent des cercles centrés en des points échantillonnés sur le contour et calculent l'aire de la proportion de chaque cercle qui est en intersection avec la région interne de la feuille à plusieurs échelles. Ainsi, des histogrammes de courbure sont construits pour chaque image.

Les courbures locales sont approximées par des triangles dans [ARKF07, AKF08, ERAK⁺05].

Alajlan et al. [ARKF07] représentent le contour des objets par des triangles à différentes échelles dans leur approche Triangle Area Representation (TAR). Chaque triangle est décrit

par son aire signée. Selon le signe de l'aire, le triangle décrit une concavité (aire positive), une convexité (aire négative) ou bien une ligne droite (aire nulle). Pour réduire la sensibilité

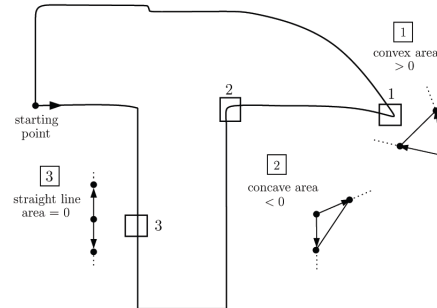


FIGURE 2.11 – Description Triangle Area Representation (TAR). Figure extraite de [AKF08]

au bruit sur le contour des descripteurs TAR et CSS, El Rube et al. [ERAK⁺05] proposent l'approche Multiscale Triangle Area Representation (MTAR), où ils appliquent la transformée d'ondelette à différents niveaux d'échelle avant de calculer le descripteur TAR. Les auteurs montrent que cette approche a obtenu une meilleure précision que la méthode CSS sur la base MPEG-7 [LLE00].

Alajlan et al. [AKF08] introduisent le Curvature Tree (CT) pour la représentation jointe de la forme et de la topologie d'un objet. Ils étudient des relations d'inclusion des objets appartenant à de différentes images. La première étape consiste à établir un graphe permettant une représentation hiérarchique de chaque image. Puis, chaque objet figurant dans l'image est décrit avec le descripteur TAR. Cette approche a montré son efficacité sur la base MPEG-7 [LLE00] et sur une base d'images médicales contenant des images synthétiques et des images réelles.

Descripteurs CCD

Les méthodes se basant sur des calculs de distances ou/et d'angles, effectués sur le contour externe de la forme ou par rapport à son barycentre, sont regroupées dans ce paragraphe. Un descripteur parmi les plus utilisés est le Centroid Contour Distance (CCD) qui calcule une séquence de distances entre le barycentre de la forme et les points du contour externe. Une fois le sens de parcours des points du contour choisi, la forme est représentée par une fonction 1D.

Cette représentation est intuitive et facile à calculer. De plus, elle est invariante par rapport à la translation et à l'échelle si le vecteur de distances est normalisé. Par contre, le

descripteur CCD n'est pas invariant par rapport à la rotation car le point de départ pour construire le vecteur de distances doit être fixe. Pour résoudre ce problème, il est possible de chercher le décalage optimal qui permet de trouver une distance minimale entre deux vecteurs mais ceci s'avère très coûteux en temps de calcul, en particulier si la taille du vecteur est importante.

Wang et al. [WCFW00] proposent une méthode efficace pour assurer l'invariance à la rotation du descripteur CCD tout en évitant de choisir un point de départ. Les auteurs calculent initialement des squelettes sur les images binaires. Ils considèrent ensuite les points extrêmes (ending points) de chaque squelette (cf. Figure 2.12(a)) et repèrent les points du contour les plus proches de ces points (cf. Figure 2.12(b)). Le décalage (shifting) du vecteur se fait donc directement pour coïncider avec les points du contour détectés. Cette méthode réduit considérablement le temps de calcul et assure l'invariance par rapport à la rotation.

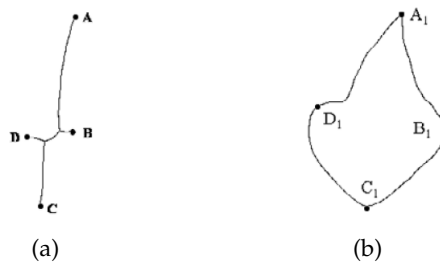
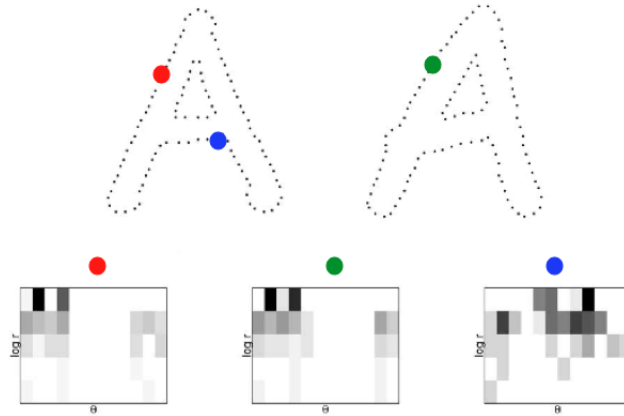


FIGURE 2.12 – Méthode de détection des points de référence. (a) squelette et les points extrêmes (b) les points du contour les plus proches des points extrêmes. Figure extraite de [WCFW00]

Dimension fractale

La dimension fractale est issue de la géométrie fractale. Elle diffère par rapport à la géométrie euclidienne par la définition de la dimension. En géométrie euclidienne, la dimension est exprimée par un entier naturel (0 : point, 1 : droite, 2 : plan, 3 : espace, etc.) alors que la géométrie fractale supporte des valeurs de dimension intermédiaires. Cela a pour but de représenter les irrégularités, trop importantes pour être décrites efficacement par la géométrie euclidienne. L'application de la technique de la dimension fractale aux images vise à analyser la complexité du contour [BB10, TFdFC04] ou de la texture en terme de son irrégularité [BB09, MCGB13].

Bruno et al. [BdFd08] comparent deux techniques de dimension fractale pour décrire les contours des feuilles : la méthode *box-counting* et la méthode de Minkowski multi-échelle et démontrent que la deuxième méthode est plus efficace pour caractériser l'espèce.

FIGURE 2.13 – La description *Shape Context*. Figure extraite de [BMP02]

Contexte de forme

Le descripteur de forme *Shape Context* (ou contexte de forme) a été proposé par Belongie et al. [BMP02]. L'idée de ce descripteur est de considérer un ensemble de N points uniformément distribués sur le contour $\mathcal{C} = \{p_1, \dots, p_i, \dots, p_N\}$ et de représenter chaque point p_i par la distribution spatiale des autres points vus de ce point (cf. Figure 3.15). La distance euclidienne (cf. Figure 2.14(a)) et l'angle sont calculés entre chaque paire de points du contour. Les distances et les directions sont quantifiées respectivement en N_D et N_A intervalles dans un repère log-polaire. Le descripteur *Shape Context* est stocké dans un vecteur 2D de dimension $N_D \times N_A$ où chaque case contient le nombre de points du contour appartenant au *bin* correspondant dans la représentation log-polaire. Ainsi, deux points ayant une configuration spatiale proche seront représentés par des histogrammes similaires. (cf. Figure 3.15). Deux travaux [LJ07, ZHW12] proposent une amélioration du descripteur *Shape Context* en modifiant la distance euclidienne utilisée par [BMP02].

Ling et Jacobs [LJ07] introduisent le *Inner Distance Shape Context* où une distance interne (*inner-distance*) remplace la distance euclidienne classique. La distance interne d'une paire de points du contour (p, q) est définie comme la longueur du chemin le plus court intérieur à la forme et joignant p et q (voir Figure 2.14(b)).

Zhi et al. [ZHW12] proposent de représenter la distance entre deux points du contour p et q par la longueur de la plus petite portion du contour entre ces deux points (cf. Figure 2.14(c)). Ils améliorent les résultats obtenus par l'approche *Inner Shape Context* sur les bases ICL et Swedish [Söd01].

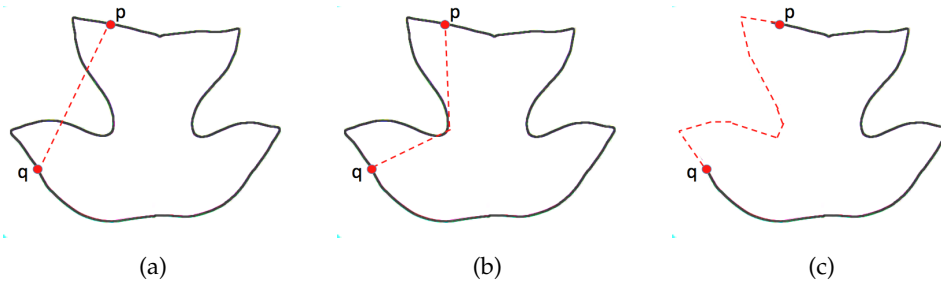


FIGURE 2.14 – Distances proposées entre deux points du contour p et q dans une description Shape Context (a) distance euclidienne [BMP02] (b) distance interne à la forme [LJ07] (c) longueur de l'arc [ZHW12]

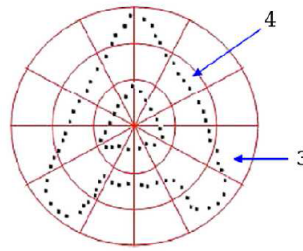


FIGURE 2.15 – Description CPDH (Contour Points Distribution Histogram). Figure extraite de [SW11]

Les trois travaux cités ci-dessus souffrent d'un problème commun qui est le coût de calcul et de comparaison d'un grand nombre d'histogrammes Shape Context. Certains travaux proposent de réduire ce nombre et d'optimiser la description avec des techniques diverses [XHS08, WFB⁺14, SW11].

Shu et Wu [SW11] introduisent le *Contour Points Distribution Histogram* (CPDH) où ils représentent la distribution des points du contour par rapport au barycentre de la forme (cf. Figure 2.15). Ils utilisent la même représentation log-polaire introduite par Belongie et al. [BMP02]. Le descripteur CPDH peut être vu comme un descripteur *Shape Context* global car il représente le contour d'une forme par rapport à son barycentre. Cette approche a été testée sur de multiples bases de formes génériques (MPEG-7 [LLE00], Kimia [SCTK98, SKK04]) et a obtenu un meilleur score que les approches Shape Context [BMP02] et IDSC [LJ07]. Cependant, dans le cas de description foliaire, la description sommaire de la forme peut être insuffisante pour différencier des espèces végétales présentant une forte similarité inter-classe.

Pour apporter plus de précision à cette description tout en réduisant le nombre d'histogrammes calculés, Xie et al. [XHS08] proposent le *Skeletal Shape Context* (SSC) où ils calculent un ensemble optimal de points du contour à partir du squelette de la forme (cf.

Figure 2.16). Les points optimaux sont généralement localisés sur les zones d'extrema de courbure de la forme. Cette méthode obtient un score de classification supérieur à 89% sur un sous-ensemble de la base Smithsonian contenant 912 images de feuilles uniformément réparties en 38 espèces.

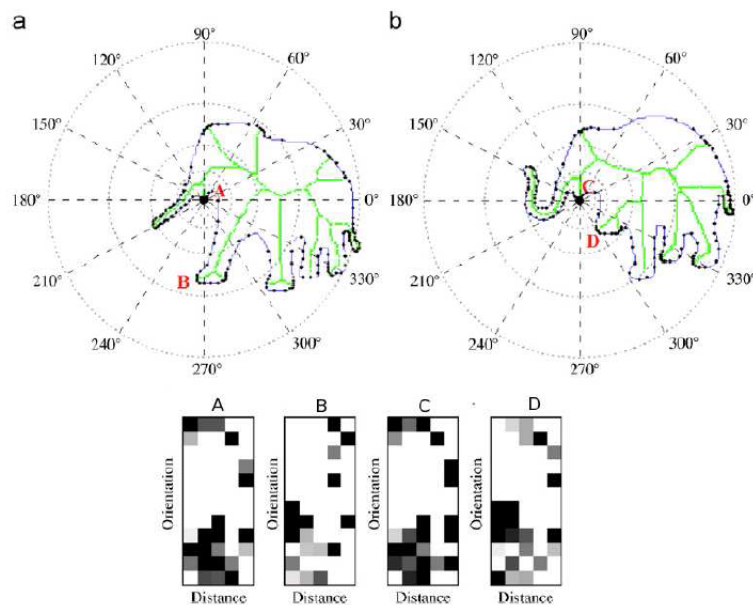
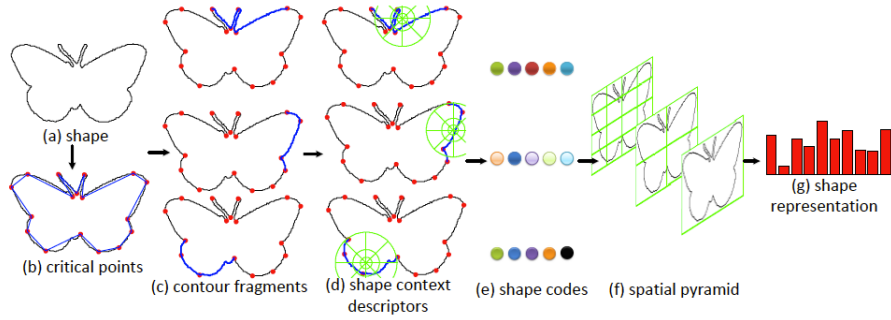


FIGURE 2.16 – Approche SSC (Skeletal Shape Context). **a** et **b** sont deux formes d'un éléphant. Les quatre histogrammes correspondent aux points de terminaison A et B (de la première squelette) et C et D (de la deuxième squelette). Notons la similarité des histogrammes des points de terminaison (A,C) et (B,D) correspondants des deux formes. Figure extraite de [XHS08]

Plus récemment, Wang et al. [WFB⁺14] ont proposé une approche appelée "sac de fragments du contour" (BCF) (cf. Figure 2.17). Cette méthode consiste à décomposer le contour en fragments saillants par la méthode DCE (Discrete Contour Evolution). Chaque fragment est ensuite décrit avec le descripteur *Shape Context* pour générer le *shape code* qui représente tout le contour. Finalement, les auteurs utilisent le *spatial pyramid matching* (SPM) pour inclure la notion de distribution spatiale des fragments du contour. Ils obtiennent de très bons résultats sur des bases de forme 2D génériques et sur la base d'image de feuilles Swedish [Söd01].

FIGURE 2.17 – L'approche *Bag of Contour Fragments*. Figure extraite de [WFB⁺14]

Descripteurs statistiques

Moments invariants

Les moments invariants sont calculés à l'intérieur d'une forme ou à partir de son contour externe. Ils vérifient un certain nombre de propriétés d'invariance, ce qui explique leur large utilisation dans la littérature. Parmi les moments les plus populaires, citons les moments de Hu [Hu62] et les moments de Zernike [Tea80].

Rappelons tout d'abord la définition statistique d'un moment. Étant donné X une variable aléatoire, le moment d'ordre k est défini en théorie de probabilité et en statistique comme suit :

$$m_k = E(X^k) \quad (2.14)$$

où $E(X^k)$ désigne l'espérance mathématique de la variable aléatoire X^k . Nous pouvons donc déduire que le moment d'ordre 1 associé à une variable aléatoire est égal à son espérance mathématique.

Les moments de Hu [Hu62] sont invariants en rotation, en translation et au changement d'échelle. Si R est une région et $p(x_p, y_p)$ un pixel de R , le moment d'ordre $i+j$ de R est donné par l'équation suivante :

$$m_{ij}(R) = \sum_{p(x,y) \in R} x_p^i \cdot y_p^j \quad (2.15)$$

Les moments de Hu sont définis comme suit :

$$\mu_1 = m_{20} + m_{02}$$

$$\mu_2 = (m_{20} - m_{02})^2 + 4m_{11}^2$$

$$\mu_3 = (m_{30} - 3m_{12})^2 + (m_{03} - 3m_{21})^2$$

$$\mu_4 = (m_{30} + m_{12})^2 + (m_{03} + m_{21})^2$$

$$\mu_5 = (m_{30} - 3m_{12})(m_{30} + m_{12})((m_{30} + m_{12})^2 - 3(m_{03} + m_{21})^2) + (m_{03} - 3m_{21})(m_{03} + m_{21})((m_{03} + m_{21})^2 - 3(m_{30} + m_{12})^2)$$

$$\mu_6 = (m_{20} - m_{02})((m_{30} + m_{12})^2 - (m_{03} + m_{21})^2) + 4m_{11}(m_{30} + m_{12})(m_{03} + m_{21})$$

$$\mu_7 = (m_{30} - 3m_{12})(m_{03} + m_{21})((m_{03} + m_{21})^2 - 3(m_{30} + m_{12})^2) - (m_{03} - 3m_{21})(m_{30} + m_{12})((m_{30} + m_{12})^2 - 3(m_{03} + m_{21})^2)$$

Les moments de Hu ont été combinés avec des descripteurs morphologiques pour décrire la forme des feuilles dans [DWZ07].

D'autres moments ont été utilisés : Zhenjiang [Zhe00] a démontré l'utilité des moments de Zernike pour la classification des variétés de roses. Les images ont été prises sur un fond uniforme et la classification est basée sur la comparaison de leurs formes. Les auteurs utilisent les moments pour décrire la rondeur (*roundness*) d'une rose.

Wang et al. [WHD⁺08] proposent une méthode basée sur les moments pour la classification des images de plantes avec un fond non uniforme. Une segmentation par la ligne de partage des eaux (LPE) contrôlée par des marqueurs est initialement effectuée afin d'extraire les régions d'intérêt. La description des feuilles est ensuite réalisée avec sept moments de Hu et douze moments de Zernike.

Une étude comparative de trois types de moments a été réalisée dans [ZSM11] afin d'évaluer leur efficacité dans un système d'identification de feuilles d'arbres. Cette comparaison inclut les moments invariants de Zernike, les moments invariants de Legendre et les moments invariants de Tchebichef. La meilleure performance est obtenue avec les moments de Tchebichef en réalisant 100% de bonne classification. Ce taux est à confirmer vue la taille de la base sur laquelle la comparaison a été effectuée. En effet, la base de test est privée et comporte 100 images de feuilles réparties en 10 espèces. De plus, les caractéristiques de cette base en terme de variabilité-inter classe et variabilité intra-classe ne sont pas fournies.

Graphes complexes

Les graphes complexes constituent une représentation statistique du contour souvent utilisée pour l'analyse des formes [BCB09, BB09, CFB11].

Backes et al. [BB09, BB10] représentent le contour par des graphes complexes où les nœuds sont des points du contour et les arrêtes sont les distances euclidiennes les reliant. Seule-

ment les distances inférieures à un seuil sont considérées. Casanova et al. [CFB11] appliquent les graphes complexes à l'identification des feuilles d'arbres lors de la tâche d'identification de plantes ImageCLEF 2011 [GBJ⁺11]. Ils adoptent la même représentation dans [BCB09] et obtiennent le meilleur score d'identification sur des photographies de feuilles et le meilleur score moyen considérant les trois catégories d'images de feuilles (scan, pseudo-scan et photographies).

Descripteurs Fourier

Certaines approches utilisent la transformée de Fourier pour décrire la forme des feuilles [KG82, RSH96, NMJS06, KNSS11b].

La transformée de Fourier 1D est une opération appliquée à un signal dans le domaine fréquentiel. Elle permet de représenter en fréquence (développement sur une base d'exponentielles), des signaux qui ne sont pas périodiques. Par analogie, on trouve les séries de Fourier pour les fonctions périodiques (développement sur la base de fonctions sinusoïdales). Une fonction non périodique peut être considérée comme une fonction dont la période est infinie. Ce passage à l'infini nous fait passer des séries aux intégrales.

La transformée de Fourier 1D $TF(f)$ d'une fonction continue f est définie dans le domaine fréquentiel par :

$$TF_{1D}(f) = \int_{-\infty}^{+\infty} f(x)e^{-i2\pi fx} dx \quad (2.16)$$

La transformée de Fourier d'une fonction est donc la généralisation au cas non périodique du calcul des coefficients de Fourier d'une fonction périodique.

Dans le cas d'une image (signal 2D), les basses fréquences représentent les surfaces homogènes caractérisées par un changement d'intensité faible ou progressif. Les hautes fréquences représentent des changements brusques d'intensité comme les contours ou le bruit. L'équation 2.16 peut s'étendre aux signaux 2D comme suit :

$$TF_{2D}(f) = F(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)e^{-i2\pi f(u.x+v.y)} dx dy \quad (2.17)$$

où x et y représentent les coordonnées spatiales et u et v sont les coordonnées spectrales.

Kuhl et Giardina [KG82] introduisent le Fourier elliptique (Elliptic Fourier EF) pour la classification des contours fermés. Cette méthode utilise les coefficients de Fourier nor-

malisés par rapport à des propriétés elliptiques et assure l'invariance par rapport aux translations, rotations et aux changements d'échelle. La description de la forme par un Fourier elliptique s'est avérée efficace pour la modélisation des pétales de fleurs [YION04] et pour l'analyse de la forme foliaire dans un contexte d'identification d'espèces de plantes [NMJS06].

Dans [KNSS11b], les auteurs présentent une approche d'identification de plantes combinant la couleur, la nervation et la forme. Cette dernière est décrite avec un Fourier polaire (Polar Fourier) où les coordonnées polaires remplacent les coordonnées spatiales.

2.3 DESCRIPTION DE LA NERVATION

Dans cette section, nous nous intéressons à un caractère particulier de la feuille : sa nervation. Des méthodes d'extraction et de description de nervures ont été proposées dans la littérature. Nous commençons par présenter la définition botanique de la nervation d'une feuille dans le paragraphe 2.3.1. Les paragraphes 2.3.2 et 2.3.3 présentent respectivement les travaux réalisés sur l'extraction et la description des nervures.

2.3.1 Nervation en botanique

La nervation désigne en botanique "l'ensemble et la disposition des nervures d'une feuille" ce qui nous amène à définir une nervure. Elle est définie par "une ligne saillante constituée de tissus conducteurs de certains organes (feuilles, pétales, sépales) et dans laquelle se trouve transportée la sève". La figure 2.18 montre la structure d'un réseau de nervure d'une feuille. Nous distinguons la nervure principale (ou encore la nervure médiane) à partir de laquelle sortent les nervures secondaires et puis les réseaux tertiaires. Les plus petites structures de nervures sont appelées micro-nervures ou fabrique de nervures. Plusieurs types de nervations sont également définis en botanique (voir Figure 2.19).

2.3.2 Extraction du réseau de nervures

Un grand nombre de méthodes d'extraction de nervures a été proposé dans la littérature. Elles peuvent être classées comme suit :

- les méthodes utilisant des détecteurs de contours génériques

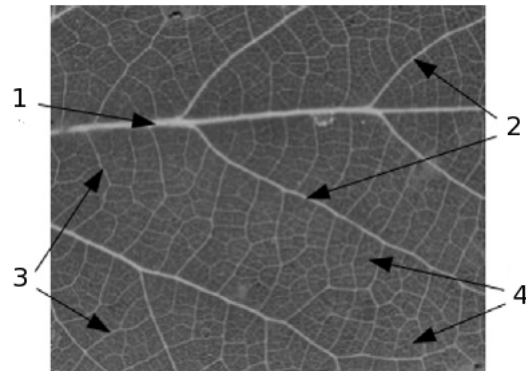


FIGURE 2.18 – Structure des nervures. 1 : nervure principale 2 : nervures secondaires 3 : nervures tertiaires 4 : micro-nervures ou fabrique de nervures. Figure de [CCC⁺12]



FIGURE 2.19 – Types de nervations. De Gauche à droite : dichotomique, alternée, pennée, palmée et curvinervée. Figure de [CCC⁺12]

- les méthodes de seuillage
- les méthodes morphologiques
- les méthodes de classification

Dans la première catégorie d'approches, les détecteurs de bord génériques utilisés peuvent être Sobel, Canny, Perwitt, *etc.* Ces méthodes utilisent la variation des niveaux d'intensité sur les images en niveaux de gris, ce qui nécessite que la couleur des nervures soit contrastée par rapport à la couleur interne du limbe. Cependant, cette hypothèse n'est pas toujours vérifiée sur les feuilles (voir image à gauche dans la figure 2.19). De plus, un détecteur de bord générique capte chaque variation importante en niveau de gris de la feuille et génère par conséquent beaucoup de bruit. Le bruit peut être dû aux changements de luminosité, à des éléments de texture ou encore à la présence de petites régions de la feuille dont la couleur est différente (tâches, trous, *etc.*).

Park et al. [PHN06, PHN08] utilisent le détecteur de bord Canny pour l'extraction des nervures.

Les méthodes de seuillage cherchent à trouver un seuil optimal séparant le réseau de nervures du reste des éléments de la feuille. Ces méthodes visent à éliminer toutes les structures n'appartenant pas aux nervures. La difficulté principale réside dans le choix d'un seuil global. En effet, il est très difficile de trouver un seuil optimal pour une image, et c'est encore plus le cas si on considère un ensemble d'images. Afin d'éviter ce problème, Bruno et al. [BdFd08] décomposent le limbe en petites régions et calculent un seuil pour chacune d'entre elles. Il s'agit d'un seuillage adaptatif.

Les méthodes morphologiques [KNSS11b, EM12, SGG10] calculent un ensemble de filtres morphologiques afin d'éliminer le bruit et ne garder que les structures qui correspondent aux nervures. En morphologie mathématique, tous les filtres reposent sur un élément structurant (SE). C'est un voisinage défini par un ensemble de pixels caractérisé par une taille et une forme : linéaire, carré, disque, *etc.*

Enficiaud et Mouine [EM12] proposent une approche morphologique pour extraire les nervures de feuilles de vigne. Une transformation *Chapeau haut de forme* est calculée afin d'extraire le contraste local de la feuille. Des ouvertures directionnelles (avec des éléments structurants linéaires) sont ensuite calculées sur l'image résultante de cette transformation suivant différentes directions.

Abdulkadir et al. [KNSS11b] appliquent des ouvertures avec un élément structurant circulaire sur l'image en niveaux de gris. Ils répètent quatre fois cette opération en faisant varier le rayon de l'élément structurant et soustraient à chaque fois l'image originale de l'image résultante de l'ouverture. Le résultat de cette approche est donné par la Figure 2.20. L'analyse du résultat d'extraction illustre deux phénomènes classiques :

- La discontinuité du réseau de nervures : elle est due généralement à un faible contraste sur des parties des nervures mais peut être également expliquée par le choix d'un élément structurant non adéquat.
- La présence des faux positifs : c'est un problème commun à la majorité des méthodes d'extraction de nervures. Dans le cas de [KNSS11b], ceci est également dû au choix de l'élément structurant circulaire. La forme circulaire n'est pas adaptée à la structure linéaire du réseau de nervures et l'élimination des pixels de bruit n'est pas assez efficace.

Les méthodes de classification sont appliquées sur tous les pixels du limbe et attribuent une étiquette à chaque pixel déterminant son appartenance au réseau de nervures. En effet, deux classes sont définies : nervure ou non-nervure. Dans ce contexte, Fu et Chi [FC06]

utilisent une méthode de classification basée sur un réseau de neurones pour décider si un pixel p appartient aux nervures. L'apprentissage est effectué sur dix valeurs calculées sur une fenêtre centrée en p . Les quatre premières valeurs sont les valeurs du gradient en p suivant quatre directions et les six valeurs restantes visent à décrire l'intensité du pixel p en fonction de son voisinage.

Cope et al. [CRBW10a] proposent une méthode d'extraction de nervures utilisant des classifieurs successifs. L'apprentissage est effectué sur neuf valeurs similaires aux valeurs calculées dans [FC06] sur un pixel p . Les classifieurs successifs basés sur l'algorithme génétique décident ensuite si le pixel p appartient au réseau de nervures. Les auteurs montrent que leur approche extrait les nervures principales et secondaires avec une grande précision tout en minimisant le bruit (cf. Figure 2.21). Cette approche a été comparée avec une autre méthode qui utilise l'algorithme de colonies de fourmis pour l'extraction des nervures. Il s'est avéré que la méthode d'extraction, utilisant les classifieurs successifs, permet un meilleur filtrage du bruit. Quant à la méthode basée sur les colonies de fourmis, elle assure une meilleure continuité des nervures. Ils concluent qu'une combinaison entre les deux méthodes pourrait améliorer la qualité du réseau de nervures extrait.

2.3.3 Description du réseau de nervures

Une fois le réseau de nervures extrait, il faut le décrire.

Park et al. [PHN08] extraient des points de courbure maximale sur les nervures. Ces points sont considérés comme des points caractéristiques des nervures. Les auteurs estiment leur densité de distribution. Ils montrent que cette information améliore le résultat d'identification obtenu avec seulement des descripteurs de forme.

Nam et al. [NHK08] décrivent le réseau par un graphe $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ construit à partir de deux ensembles : un ensemble de noeuds \mathcal{V} où chaque noeud est un point caractéristique



FIGURE 2.20 – Résultat d'extraction de nervures dans [KNSS11b]. Figure extraite de [KNSS11b]



FIGURE 2.21 – Résultat d'extraction de nervures dans [CRBW10a]. Figure extraite de [CRBW10a]

et un ensemble \mathcal{E} décrivant les arêtes les reliant (cf. Figure 2.22). Les deux méthodes ci-

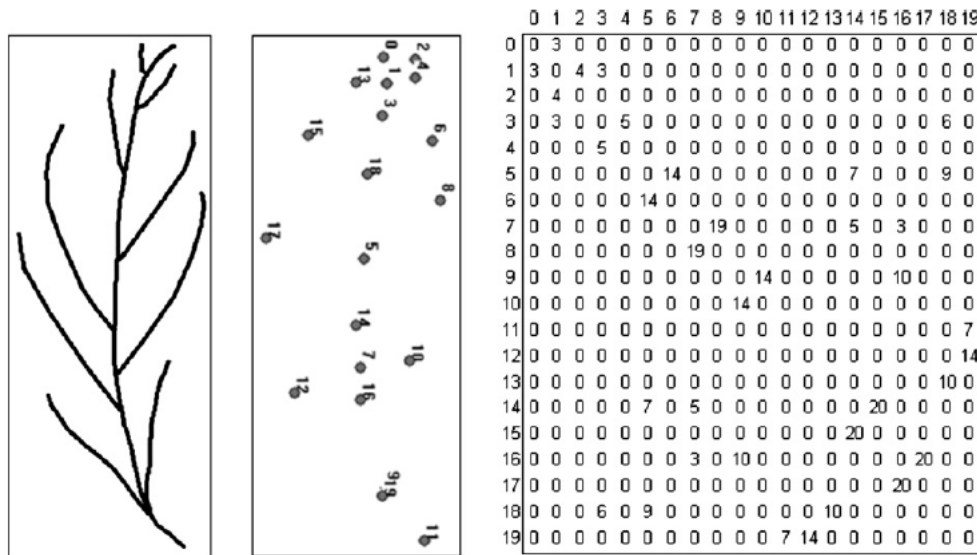


FIGURE 2.22 – Points caractéristiques des nervures et matrice d'adjacence dans [NHK08]. Figure extraite de [NHK08]

tées ci-dessus [PHN08, NHK08] supposent que l'étape d'extraction produit un réseau de nervures totalement connecté sinon les notions de points d'extrémité (ending points) et de points de ramification perdent leur sens. Or en fait, les discontinuités des nervures sont très fréquentes ce qui explique le fait que certaines approches ont recours à une phase de reconstruction [EM12].

Dans [KNSS11b], les auteurs calculent le rapport entre la surface des nervures et la surface de la feuille en terme de nombre de pixels. Ils obtiennent quatre descripteurs à partir de quatre images de nervures différentes, calculées à l'aide des éléments structurants de diffé-

rentes tailles.

La technique de dimension fractale a été également utilisée dans [BdFd08] pour la description des nervures.

2.4 DESCRIPTION DE LA TEXTURE

Il n'existe pas une définition conventionnelle de la texture dans une image. La texture décrit généralement les éléments répétitifs dans une image. Pour les feuilles d'arbres, une texture fait référence aux micro-nervures et aux petits motifs qui se répètent. Cinq méthodes sont essentiellement utilisées pour décrire et caractériser la texture d'une feuille. Les méthodes les plus courantes sont basées sur la technique LBP ou sur les ondelettes.

2.4.1 Descripteurs LBP

Le descripteur de texture LBP (Local Binary Pattern) a été introduit par Ojala et al. [OPH96]. Il a été largement utilisé dans diverses applications, essentiellement pour la reconnaissance faciale. Le descripteur LBP est la représentation des structures spatiales locales en comparant la valeur d'intensité de chaque pixel d'une image avec ses voisins. L'opérateur LBP appliqué à un pixel p est défini comme suit :

$$LBP(P, R) = \sum_{p=0}^{P-1} s(q_p - q_c) 2^p \quad (2.18)$$

où P désigne le nombre de pixels dans le voisinage défini par le rayon R , q est le niveau de gris du pixel et la fonction $s(x) = 1$ si $x \geq 0$ sinon $s(x) = 0$.

Deux types de codes LBP sont distingués : uniformes et non uniformes. Un code LBP est dit uniforme s'il contient au plus deux transitions de 0 à 1 (ou l'inverse) quand il est parcouru de manière circulaire ; sinon le code est non uniforme. Considérant un voisinage de pixels, un code uniforme signifie qu'il y a une variation (croissante ou décroissante) uniforme par rapport à l'intensité du pixel central. Le code non uniforme, au contraire, décrit une suite de variations croissantes et décroissantes par rapport à la valeur du pixel central. Il est souvent représentatif du bruit.

En 2002, Ojala et al. [OPM02] proposent une approche multi-résolution du LBP qui est invariante par rapport à la rotation. A partir de cette méthode, un grand nombre de variantes du LBP a été proposé. Les améliorations essentielles concernent la représentation du voisinage, la représentation multi-échelle et multi-résolution, le seuillage effectué pour avoir le

code binaire du voisinage, *etc.*

Différentes définitions du voisinage ont été proposées : Centralized BP [FW08, HPS09], Three-patch et Four-patch LBP (TPLBP et FPLBP) [WHT08], Completed LBP (CLBP) [GZZ10], Fuzzy LBP (FLBP) [IKM08] et Local Derivative Pattern (LDP) [ZGZL10].

Une étude comparative de toutes ces variantes du LBP est effectuée par Nanni et al. [NLB12] sur six bases de test dont une base de feuilles. Ils en proposent une nouvelle variante (LQP).

2.4.2 Ondelettes

Une ondelette est définie en traitement du signal par une fonction oscillatoire qui est similaire à la fonction de Fourier. La fonction de Gabor est définie par l'association d'une gaussienne et d'une sinusoïde orientée (cf. Figure 2.23). L'application de cette fonction aux images donne la fonction suivante définie dans le domaine spatial :

$$G(x, y, \theta, f) = e^{-\frac{1}{2}(\frac{x_{\theta}^2}{\sigma_x^2} + \frac{y_{\theta}^2}{\sigma_y^2})} \cos(2\pi f x_{\theta}) \quad (2.19)$$

avec

$$x_{\theta} = x \cos \theta + y \sin \theta$$

$$y_{\theta} = y \cos \theta - x \sin \theta$$

où θ est l'orientation de la sinusoïde, f la fréquence de la sinusoïde et σ_x et σ_y représentent respectivement l'écart type de la gaussienne selon l'axe des abscisses et l'axes des ordonnées.

Les filtres de Gabor sont utilisés pour l'analyse de la texture des feuilles dans beaucoup

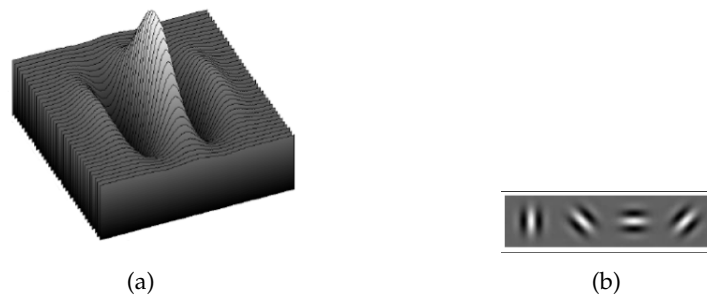


FIGURE 2.23 – Transformée de Gabor. (a) représentation 3D de la fonction de Gabor normalisée entre 0 et 255. (b) banc de filtres Gabor composé de 4 filtres avec différentes orientations.

d'approches [LZWM08, CdMSJB09, CRBW10b, KYU11]. D'autres auteurs proposent d'uti-



FIGURE 2.24 – variation de la couleur des feuilles. De Gauche à droite : un ginkgo, un érable en automne, un liquidambar en automne et un chêne rouge en automne. images dans <http://classe5.voila.net/sciences/automne/automne.htm>

liser les ondelettes de HAAR [ASF⁺11].

2.4.3 Autres descripteurs de texture

D'autres types d'approches ont été présentés. Certains d'entre eux sont basés sur la dimension fractale [BB09, MCGB13] et d'autres méthodes utilisent les matrices de co-occurrence calculées à partir des images couleur [KYU11]. Dans le cadre de la campagne d'évaluation ImageCLEF 2011, le descripteur Fourier associé à un voisinage des points d'intérêt a été utilisé par la meilleure approche pour les feuilles scannées.

2.5 DESCRIPTION DE LA COULEUR DES FEUILLES

2.5.1 Couleur des feuilles en botanique

La couleur de la feuille dépend de la présence d'un pigment appelé *chlorophylle*. Il a pour rôle la sécrétion des sucres nécessaires à la nutrition de la plante à partir de l'énergie provenant des rayons solaires. La chlorophylle est renouvelée en permanence durant la période du printemps et de l'été. La présence intense de chlorophylle est à l'origine de la couleur verte de la feuille. A la fin de l'été, le niveau de chlorophylle se met à décroître agissant ainsi directement sur la couleur de la feuille qui passe alors par différentes teintes ; jaunes, orangées et rouges à mesure que les pigments se dégradent. Ceci est illustré dans la Figure 2.24.

2.5.2 Méthodes de description de couleur des feuilles

Les premiers travaux utilisant la couleur des feuilles [FGU91, WMVBM95] avaient pour but la détection des mauvaises herbes.

Franz et al. [FGU91] calculent trois indices colorimétriques : la moyenne, la variance et

un indice d'asymétrie pour les canaux rouge et bleu de l'espace de couleurs RVB et un coefficient de réflectance infrarouge sur des images contenant des mauvaises herbes.

Woebbecke et al. [WMVBM95] utilisent quatre valeurs dans l'espace de couleurs RVB pour séparer les mauvaises herbes du fond sur des photographies. Ils concluent que l'utilisation de la couleur peut s'avérer utile pour la segmentation des mauvaises herbes mais reste inefficace pour la classification des différentes espèces.

El Fakih et al. [EFZP00] présentent une méthode pour différencier le blé et la plante de soja des mauvaises herbes sur des images avec un fond naturel. Ils introduisent quatre indices de couleur calculés sur les trois canaux de l'espace RVB. Ces indices représentent la luminance et le contraste de chaque pixel appartenant aux feuilles. Une méthode statistique est utilisée ensuite pour trouver la meilleure combinaison entre les indices calculés.

La couleur des feuilles n'est pas suffisamment discriminante dans un système d'identification de plantes. En fait, elle dépend étroitement des saisons, des régions géographiques et de l'exposition au soleil comme nous l'avons expliqué dans la paragraphe 2.5.1. Cependant, la couleur peut améliorer l'identification lorsqu'elle est combinée avec d'autres primitives de la feuille [KYU11, KNSS11b].

Kadir et al. [KNSS11b] présentent une approche d'identification de feuilles qui combine la couleur, la forme et la nervation. Pour représenter la couleur, ils utilisent les trois indices de couleur introduits dans [FGU91].

Deux techniques de description de couleur sont utilisées dans l'approche proposée par Kebapci et al. [KYU11] : les histogrammes de couleur normalisés dans l'espace RVB (Rouge Vert Bleu) et les matrices de co-occurrence calculées dans l'espace de couleurs TSL (teinte saturation luminance).

Yanikoglu et al. [YAT12] proposent une approche de classification d'espèces de feuilles où ils combinent des descripteurs de couleur, de texture et de forme. Pour décrire la couleur, ils calculent des moments de couleur et les stockent dans un vecteur contenant vingt-sept valeurs.

2.6 COMPARAISON ET MISE EN CORRESPONDANCE

L'identification de l'espèce d'une image de feuille passe par une phase de comparaison (matching) qui consiste à mettre en correspondance les images visuellement semblables.

Ce processus suit l'un des deux schémas suivants :

- la recherche : cette technique dénommée *Image retrieval* se base sur la recherche d'images similaires et permettant d'en déduire la classe de l'image requête.
- la classification : cette technique utilise les algorithmes classiques de classification i.e. *clustering*) ou se base sur l'apprentissage automatique.

2.6.1 Recherche d'images par similarité (Image Retrieval)

C'est la technique de base utilisée par les CBIR (Content-based image retrieval) où trois phases sont nécessaires :

L'étape initiale est l'indexation qui consiste à calculer et stocker des caractéristiques visuelles de toutes les images d'une base. Elle est effectuée hors ligne, c'est-à-dire avant que l'utilisateur ne commence à soumettre des requêtes. L'indexation produit, à partir d'une base d'images, une base d'index comme le montre la Figure 2.25. Dans la base d'index, chaque image est représentée par des descripteurs de ses attributs visuels. Dans le processus de recherche d'images par similarité, l'utilisateur sélectionne une image requête dont la classe (espèce dans notre cas) est supposée inconnue. La signature de l'image requête est calculée puis elle est comparée aux signatures se trouvant dans la base d'index. Lors de cette comparaison, une distance est calculée entre chaque paire de signatures. Elle mesure la similarité visuelle entre les images.

Le processus de comparaison suit l'un des schémas suivant :

- la recherche exhaustive où chaque signature est comparée à l'ensemble des signatures restantes. Cette technique est très coûteuse en temps de calcul particulièrement si la taille de la base est importante. Elle est donc rarement utilisée par les systèmes de recherche par le contenu.
- la recherche approximative où chaque signature est comparée à un ensemble réduit de signatures, choisi de façon à contenir un grand nombre de signatures similaires. Les techniques de recherche approximative [JB08, JB11] sont généralement appliquées dans le cas d'appariement à large échelle [JB09] mais ont également démontré leur efficacité dans la recherche pertinente des images de feuilles similaires [GJY⁺11].

Après la comparaison, une liste ordonnée d'images similaires à l'image requête est établie. L'idée est d'exploiter cette liste afin d'estimer l'espèce. Une règle de décision se basant sur les k images les plus similaires ($k - nn$: k -nearest neighbors) peut être utilisée.

Les mesures de similarité utilisées dépendent du type de description choisi pour représenter la feuille. Elles sont décrites dans le paragraphe suivant.

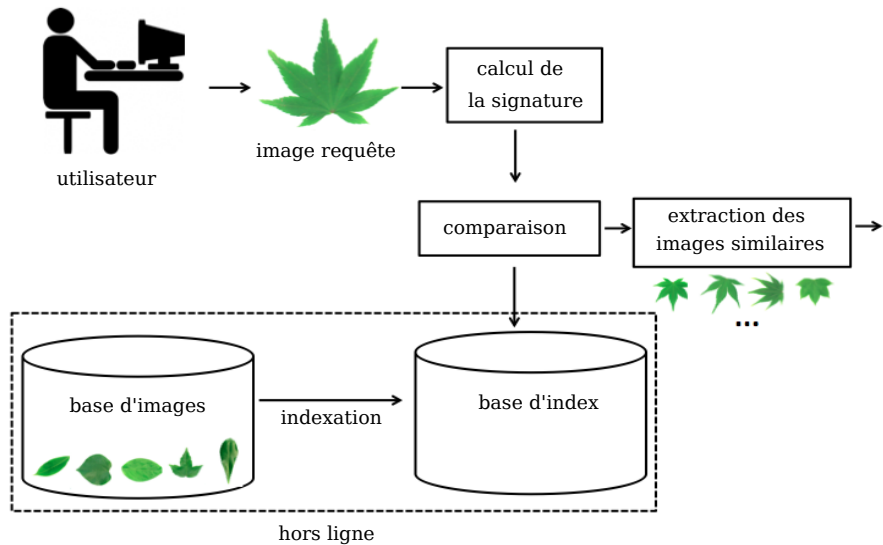


FIGURE 2.25 – Architecture d'un système de recherche d'images par le contenu

Types de similarité

L'estimation de la ressemblance entre les images est conditionnée par les primitives choisies pour les représenter. Elle est souvent exprimée par une mesure de similarité qui consiste à trouver des métriques appropriées pour modéliser le degré de proximité entre les images. Deux types de mesure de similarité sont distinguées :

- les mesures de similarité globale
- les mesures de similarité locale

Ces deux types sont discutés dans les deux paragraphes suivants.

Similarité globale

Une mesure de similarité globale permet de comparer deux images dans leur globalité. Elle suppose que chaque image est décrite par une primitive globale (dite aussi complète). Pour comparer deux images, il suffit de comparer leurs deux signatures respectives. Une mesure de similarité globale a l'avantage d'être intuitive et rapide. En revanche, elle est sensible au bruit puisqu'une légère différence au niveau visuel entre deux images peut éventuellement conduire à des signatures très différentes. Les approches suivantes :

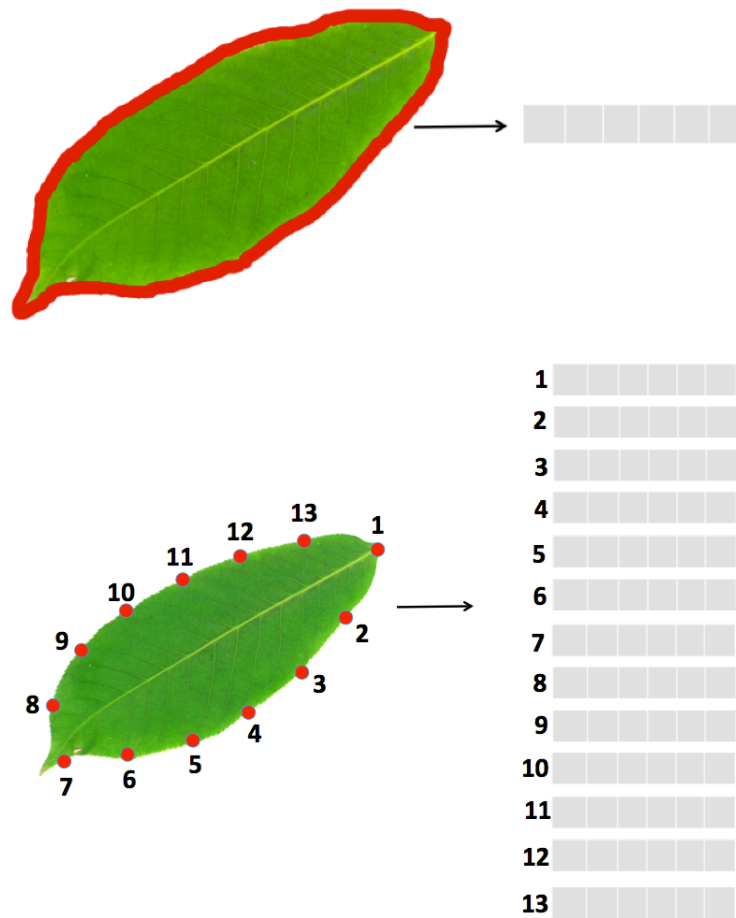


FIGURE 2.26 – Description globale et description locale du contour

- les moments invariants,
 - les descripteurs morphologiques issus des paramètres de forme,
 - la dimension fractale,
 - les histogrammes de couleur,
- utilisent une similarité globale.

Similarité locale

Dans le cas d'une mesure de similarité locale, chaque primitive décrite correspond à une partie visuelle partielle dans l'image (points d'intérêt, points du contour [MYVB12], une petite fenêtre autour d'un point d'intérêt [GJY⁺11], *etc.*). La signature d'une image est composée d'un ensemble de descripteurs 2.26. Chacun d'entre eux est associé à la primitive locale qu'il décrit. Pendant la phase de comparaison, chaque signature locale est comparée à l'ensemble des signatures locales des images de la base. D'une manière

générale, deux images sont considérées comme similaires quand elles possèdent un grand nombre de signatures locales similaires. Un exemple typique d'approches appartenant à cette catégorie sont celles qui utilisent les points d'intérêt. La mesure de similarité locale est, en général, robuste au bruit puisque une différence locale entre deux images n'a pas une grande influence sur le calcul de similarité. La similarité locale offre aussi une certaine robustesse à l'occultation partielle.

Les mesures de similarité, locales ou globales, dépendent du calcul d'une distance entre les signatures ou les vecteurs. Les distances les plus courantes utilisées dans la littérature sont : la distance euclidienne, la distance de Hamming et la distance de Hausdorff [GBVC⁺13].

2.6.2 Classification

La classification repose sur l'apprentissage automatique qui prend son origine dans le domaine de l'intelligence artificielle. Son principe consiste en une étape d'apprentissage où un système est entraîné, à partir d'un ensemble d'entrées, à prédire la classe d'une image donnée. Ceci peut s'appliquer parfaitement à un système d'identification de plantes. L'entrée du système d'apprentissage va être un ensemble d'attributs foliaires calculés au préalable, et la sortie est le résultat de prédiction de l'espèce. Parmi les méthodes d'apprentissage les plus utilisées pour l'identification des plantes, nous trouvons les machines à vecteurs de support (Support Vector Machine, SVM) [NZ08, RSBG13, RWZ12], les réseaux de neurones artificiels [ASFRRGG11, KNSS11a, ON00, SK00, UU12, WBX⁺07, ZSM11] et les forêts aléatoires [Cer13].

2.7 RÉCAPITULATION ET DISCUSSION

En se basant sur cet état de l'art et sur le tableau récapitulatif 2.2, nous pouvons souligner que :

- la majorité des méthodes d'identification de feuilles sont basées sur l'analyse de la forme. Cela reflète la grande importance de ce caractère foliaire et sa grande capacité de distinction entre les espèces. La forme est toujours retenue dans les méthodes qui combinent différents caractères de la feuille. En terme de taux d'identification, les meilleures méthodes sont dans la plupart des cas, basées sur la forme.

Approche	caractère(s) décrit(s)	descripteurs	type de classification
Du et al. [DWZ07]	forme	descripteurs géométriques + moments invariants	classification : MMC hypersphere
Wang et al. [WCF03]	forme	excentricité + centroid-contour distance (CCD) + angle code histogram (ACH)	Recherche
Caballero et Aranda [CA10]	forme	descripteurs géométriques + descripteur de courbure	Recherche
Uluturk et Ugur [UU12]	forme	descripteurs géométriques	Classification(réseau de neurones probabiliste(PNN))
Cerruti et al. [CTM ⁺ 13]	forme	descripteurs géométriques botaniques + histogrammes de courbure	classification
Ling et Jacobs [LJ07]	forme	Inner Distance Shape Context IDSC	Recherche
Alajlan et al. [ARKF07]	forme	Triangle area representation (TAR)	Recherche
EL Rube et al. [ERAK ⁺ 05]	forme	descripteurs géométriques + multiscale triangle area representation (MTAR)	Recherche
Wang et Gao [WG13]	forme	descripteurs géométriques + descripteur de courbure	Recherche
Kumar et al. [KBB ⁺ 12]	forme	Histogrammes de courbure	Recherche
Bruno et al. [BdFd08]	forme et nervation	dimension fractale	classification : linear discriminant analysis (LDA)
Lin et al. [LZWM08]	texture	filtres de Gabor + LBP	classification : SVM
Nanni et al. [NLB12]	texture	LBP	classification : SVM
Casanova et al. [CdMSJB09]	texture	ondelettes de Gabor	classification
Cope et al. [CRBW10b]	texture	filtres de Gabor	classification
Backes et Bruno [BB09]	texture	dimension fractale multi-échelle	classification :LDA
Franz et al. [FGU91]	couleur	moments de couleur	classification
Nam et al. [NHK08]	forme	Minimum Perimeter Polygons (MPP)	Recherche
	nerivation	graphe de points caractéristiques + matrice d'adjacence	
Kadir et al. [KNSS11a]	forme	Polar Fourier	Recherche
	couleur nervation	moments de couleurs rapport d'aires nervures/feuille	
Kebapci et al. [KYU11]	forme	SIFT + descripteurs botaniques + courbure	Recherche
	couleur	histogrammes de couleur + matrices de co-occurrence	
	texture	ondelettes de Gabor	

TABLE 2.2 – Etat de l'art des méthodes d'identification/classification de feuilles.

- peu d'approches se basent sur la couleur de la feuille. Cela confirme que la couleur ne peut pas être une primitive déterminante dans un système d'identification d'espèces végétales. Certaines approches récentes, proposent de la combiner avec d'autres caractères plus discriminants mais son apport reste relativement minime.
- les approches basées sur les points d'intérêt, peu exploitées, ont pourtant obtenu de bons scores dans les tâches d'identification de plantes ImageCLEF en 2011 [GBJ⁺11] et en 2013 [GJB⁺13].
- les méthodes qui combinent plusieurs caractères foliaires obtiennent généralement de bons résultats d'identification. Ceci provient certainement du choix des primitives et de leur complémentarité.

Pour conclure, l'étude des méthodes existantes nous a permis d'identifier les points forts et les limites de chaque type d'approches. Nous avons également remarqué que d'autres approches prometteuses ne sont pas assez exploitées dans la problématique d'identification d'espèces, notamment les approches locales.

2.8 CHOIX ET HYPOTHÈSES

2.8.1 Types d'images

Dans ce travail, nous nous focalisons sur la description de la feuille. Une étape préliminaire d'extraction des éléments saillants (comme le contour, la nervation, *etc.*) est nécessaire. La complexité de cette étape varie selon la nature des images traitées. Les types d'images sont liés à l'appareil (photographie, scanner, *etc.*) et au protocole d'acquisition.

Nous distinguons trois types d'images de feuilles (cf. Figure 2.27) :

- Les scans de feuilles : ce sont des images acquises via un scanner. Les scans sont généralement des images de bonne qualité où le fond est complètement homogène.
- Les pseudo-scans : cette catégorie contient les images prises par un appareil de photographie sur un fond uniforme. Ces images peuvent comporter des ombres portées, peuvent être floues ou avoir un problème de contraste.
- Les photographies : cette catégorie comporte les images prises librement sans aucune contrainte sur le protocole d'acquisition. Dans ce cas, la feuille peut se confondre avec le fond.

Ces trois types d'images sont présentés dans la figure 2.27.

Dans ce travail nous considérons les deux premiers types i.e les scans et les pseudo-scans.



FIGURE 2.27 – Catégories d'images de feuilles. De gauche à droite : feuille scannée, pseudo-scan de feuille et photographie de feuille.

2.8.2 Extraction du contour

La segmentation est une étape nécessaire pour décrire la forme de la feuille.

Les trois types d'images (scans, pseudo-scans et photographies) induisent des problèmes de segmentation différents. Plus le fond est non homogène et bruité, plus la segmentation est plus complexe. Dans le cas des photographies de feuilles, Cerruti et al. [CTM⁺13] ont proposé une méthode de segmentation semi-automatique efficace. Dans le cas des scans et des pseudo-scans, la plupart d'approches utilisent une méthode de seuillage. Néanmoins, Soares et Jacobs [SJ13] ont proposé une méthode de segmentation adaptée à ce type d'images plus robuste qu'une méthode de seuillage.

Nous avons supposé tout d'abord que la forme d'une feuille peut être décrite par un unique contour fermé. Pour isoler la feuille du fond, notre choix s'est porté sur la méthode de seuillage d'Otsu [Ots79] utilisée dans de nombreux travaux. Elle consiste à séparer l'image en deux régions en maximisant la variance entre les régions et en minimisant la variance intra-région. La formulation mathématique de la variance intra-classe est définie par la somme pondérée des variances des deux classes, donnée par l'équation suivante :

$$\sigma_w^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t) \quad (2.20)$$

Otsu a démontré que minimiser la variance intra-classe revient à maximiser la variance inter-classe. Le seuil d'Otsu est donc la valeur de t , parmi tous les seuils possibles, qui minimise $\sigma_w^2(t)$.

L'application de la méthode de seuillage d'Otsu est efficace si nous considérons la nature des images traitées et elle permet généralement d'aboutir à une segmentation correcte de la surface de la feuille. Néanmoins, certains résultats restent imparfaits. Les cas typiques sont ceux des feuilles composées où il existe un recouvrement entre les folioles ou ceux

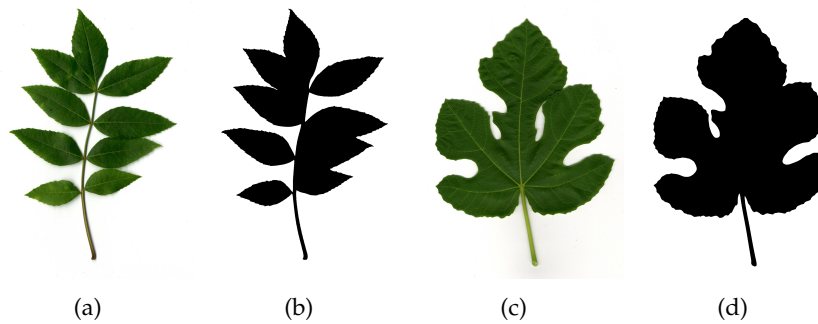


FIGURE 2.28 – Imperfections de segmentation avec la méthode de seuillage d'Otsu [Ots79] (a,b) recouvrement des folioles dans une feuille composée de l'espèce *Fraxinus angustifolia* (c,d) recouvrement des lobes d'une feuille simple lobée de l'espèce *Ficus carica*

des feuilles simples lobées où les lobes se recouvrent comme l'illustre la figure 2.28. Nous admettons ces petites imperfections du contour, inévitables dès lors que nous cherchons à avoir un seul contour fermé. Notre optique consiste à remédier à ces incorrections de contour en utilisant des descripteurs de forme robustes aux occultations partielles et aux imprécisions sur le contour.

2.8.3 Identification basée sur la recherche par similarité

Notre travail doit s'intégrer à l'outil d'identification développé dans le cadre du projet Pl@ntNet. Celui-ci utilise le moteur de recherche visuelle IKONA-Maestro [BFF⁺01] et se fonde sur une recherche par similarité.

Processus d'identification

Le processus d'identification suppose que nous disposons d'une base d'images de feuilles étiquetées. L'information indispensable qui doit figurer dans la vérité terrain est le nom de l'espèce végétale de chaque image de la base. D'autres données comme par exemple, l'individu (arbre) auquel appartient la feuille, ne sont pas nécessaires mais peuvent être exploitées.

Intéressons-nous maintenant au processus d'identification de base.

Une étape d'indexation de l'ensemble de la base d'images étiquetées s'effectue hors ligne (calcul de descripteurs + stockage dans la base d'index). La deuxième entrée est une image requête dont l'espèce est supposée inconnue. Les étapes d'identification sont les suivantes :

- calcul de la signature de l'image requête
- comparaison de la signature de l'image requête avec celles des images indexées.

- Calcul d'une liste ordonnée d'images similaires dont la similarité est décroissante
- Estimation de l'espèce de l'image requête à partir de la liste des k plus proches voisins et d'une règle de décision.

Mesure de similarité locale

La plupart des descripteurs présentés dans la suite de ce manuscrit sont des descripteurs locaux. Comme nous l'avons présenté dans le paragraphe 2.6.1, chaque image I est décrite par n descripteurs d_1, d_2, \dots, d_n . La comparaison de chaque descripteur d_i avec l'ensemble des descripteurs \mathcal{D} de la base peut se ramener au problème de la recherche des plus proches voisins dans un espace multidimensionnel de grande taille.

Afin d'accélérer ce processus, une technique de recherche dérivée de l'algorithme LSH (Locality Sensitive Hashing) est utilisée. Cet algorithme de recherche approximative apporte une solution à la recherche des plus proches voisins en grande dimension en réduisant l'espace et le coût de la recherche. L'idée est d'utiliser des fonctions de hachage "locality sensitive", c'est-à-dire telles que deux points proches ont une forte probabilité d'obtenir la même valeur de hachage et que deux points éloignés obtiennent une même valeur de hachage avec une faible probabilité. Dans le cadre de cette thèse, une variante de LSH, la méthode Posteriori Multi-Probe Locality Sensitive Hashing [JB08] implémentée dans [BFF⁺01] est utilisée.

Chapitre 3

Approches proposées

DANS ce chapitre, nous présentons trois nouvelles méthodes de description de feuilles en utilisant différents caractères. En premier lieu, une méthode de description de la forme foliaire est introduite. Elle se base sur une représentation triangulaire multi-échelle du contour. Plusieurs façons de décrire les triangles à différentes échelles sont étudiées. En deuxième lieu, nous proposons de décrire la feuille par ses points saillants. Cela nécessite une étape d'extraction des points caractéristiques de la feuille. Puis, pour décrire ces points, nous introduisons différents scénarios afin d'étudier l'impact des aspects suivants dans l'identification :

- les relations spatiales entre les points saillants
- la corrélation spatiale entre les points saillants et le contour
- les caractéristiques visuelles locales autour des point saillants

Enfin, une approche d'extraction et de description de la nervation est présentée. L'extraction des nervures est effectuée avec des outils morphologiques. Les réseaux de nervures extraits sont ensuite décrits par leurs directions principales et leurs arrangements spatiaux.

3.1 DESCRIPTION TRIANGULAIRE DE LA FORME

La forme d'une feuille est un élément caractéristique pertinent dans l'identification des plantes. Malgré les transformations que subit la feuille au cours de son cycle de développement, sa forme reste le caractère le plus intuitif pour guider le processus d'identification. La caractérisation de la forme d'une feuille représente deux véritables défis : le premier consiste à extraire correctement le contour et le deuxième consiste à l'exploiter et le décrire de la manière la plus efficace.

Le premier challenge a été évoqué dans la littérature et peut se ramener à un problème de segmentation. Des approches automatiques ou semi-automatiques ont été proposées pour résoudre ce problème [CTM⁺13, GBVC⁺13, SJ13]. Dans ce présent travail, la problématique de la segmentation a moins d'importance puisque nous nous sommes limités à des images de feuilles prises sur un fond uniforme (des scans et des pseudo-scans). C'est la raison pour laquelle nous avons choisi d'utiliser une simple méthode de seuillage [Ots79].

Nous nous focalisons sur le deuxième point qui porte sur la description de la forme foliaire. Nous avons choisi de suivre une méthode de représentation triangulaire du contour. Nous avons étudié les limites du descripteur triangulaire de forme *TAR* (Triangle Area Representation) [ARKF07] pour proposer un ensemble de descriptions alternatives. Trois nouvelles représentations sont proposées et intégrées dans une approche locale.

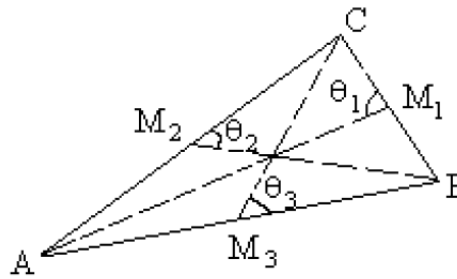


FIGURE 3.1 – Représentation *TSR*. A , B et C sont des composantes visuelles représentées par leurs centre de gravité. M_1 , M_2 et M_3 représentent les points de milieu des côtés du triangle ABC . θ_1 , θ_2 et θ_3 sont les angles aigus aux points M_1 , M_2 et M_3 . Figure extraite de [GN01]

3.1.1 Représentation triangulaire entre des objets

La représentation triangulaire a été utilisée dans la littérature dans un contexte plus général pour décrire les relations spatiales entre des entités visuelles dans une image [GN01, PG08, HGBRM10]. Ces entités peuvent représenter des objets ou des régions d'intérêt.

Guru et Nagahushan introduisent la méthode *TSR* (Triangle Spatial Relationships). Elle consiste à représenter chaque groupe de trois composantes visuelles, non colinéaires, par un triangle (cf. Figure 3.1). Supposons que L_a , L_b et L_c sont respectivement les étiquettes des composantes visuelles A , B et C , les auteurs décrivent le triangle de la figure 3.1 par un quadruplet $(L_{i1}, L_{i2}, L_{i3}, \theta)$ choisi parmi six quadruplets possibles selon des critères définis (se référer à [GN01] pour plus de détails).

La principale limite de cette méthode est qu'elle ne permet pas de représenter trois entités visuelles colinéaires dans l'image. Pour résoudre cela, Punitha et Guru [PG08] proposent une variante de *TSR* où ils rendent possible le calcul de l'angle θ dans le cas de trois composantes colinéaires.

Hoàng et al. [HGBRM10] introduisent la méthode Δ -*TSR* qui étend *TSR*. Ils l'appliquent sur des entités visuelles locales basées sur les points saillants d'un modèle *BoF* (Bag of Features). Afin de réduire la complexité de la représentation *TSR*, les auteurs suggèrent de considérer les entités visuelles à l'intérieur d'un voisinage local. Ils proposent ensuite plusieurs stratégies de représentation triangulaire suivant un point particulier, dit point pivot (cf. Figure 3.2). La description du triangle formé par les trois composantes visuelles est enrichie par l'ajout d'une information sur l'orientation des sommets et sur le facteur d'échelle. Chaque triplet est représenté par sept valeurs.

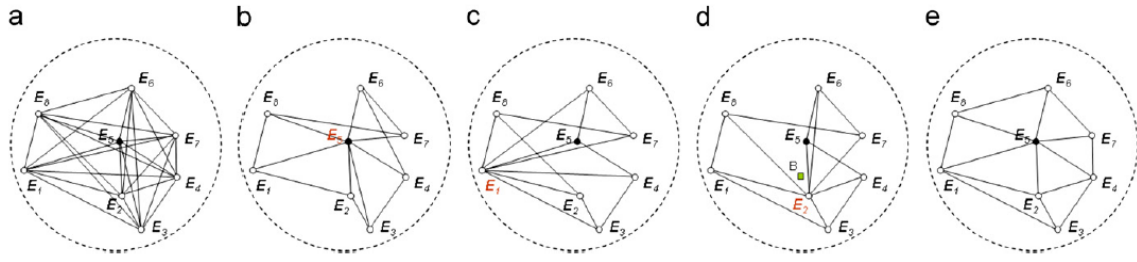


FIGURE 3.2 – Stratégies de représentation triangulaire. (a) Tous les triangles possibles (b) le point pivot est le centre du voisinage (c) le point pivot est le premier point dans la liste (ici E_1), sachant que la liste est triée (c) le point pivot est le plus proche du barycentre de l'ensemble de points (d) triangulation de Delaunay. Figure extraite de [HGBRM10]

La modélisation des relations spatiales en général est essentielle dans d'autres problèmes de reconnaissance de forme, comme par exemple, dans celui de la reconnaissance des symboles (cf. [SLW12] et [SWL14]).

Les représentations triangulaires citées ci-dessus visent à décrire les relations spatiales entre des entités visuelles dans une image. Nous nous focalisons dans la suite sur la représentation du contour externe d'un objet unique.

3.1.2 Représentation triangulaire locale du contour

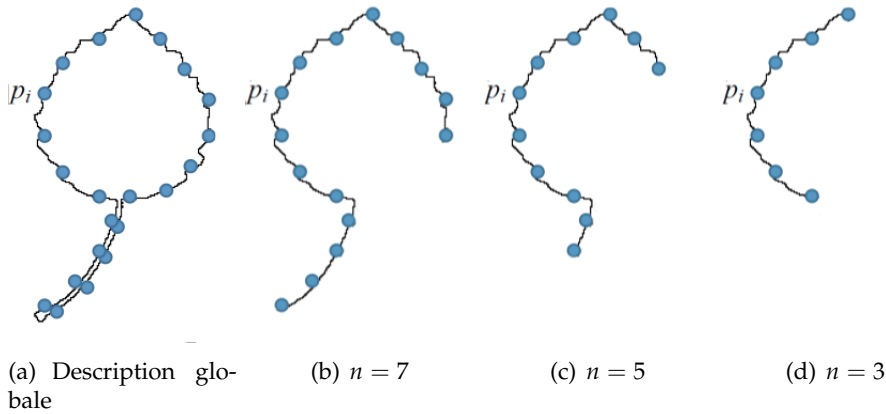


FIGURE 3.3 – Portion du contour retenue suivant le paramètre n

La forme d'une feuille est représentée par un ensemble \mathcal{P} composé de N points échantillonnés sur son contour $\mathcal{P} = \{p_1, \dots, p_N\}$. Ces points sont distribués d'une façon uniforme sur le contour et sont numérotés dans le sens des aiguilles d'une montre. Nous devons sélectionner une portion du contour $\mathcal{W} \subset \mathcal{P}$ pour décrire un point p_i et son voisinage. Cette portion du contour \mathcal{W} est décrite par un ensemble de points appartenant à \mathcal{P} et voisins de p_i . Nous définissons le paramètre $n = 1, \dots, \frac{N-1}{2}$ tel que \mathcal{W} peut s'écrire comme suit :

$$\mathcal{W} = \{p_{i-n}, \dots, p_{i-1}, p_i, p_{i+1}, \dots, p_{i+n}\} \quad (3.1)$$

Ainsi, si par exemple $n = 3$, nous allons considérer la portion du contour qui relie :

- les 3 points de \mathcal{P} voisins de p_i , en partant dans le sens contraire des aiguilles d'une montre
- le point p_i
- les 3 points de \mathcal{P} voisins de p_i , en partant dans le sens des aiguilles d'une montre

Nous pouvons alors trouver la cardinalité de \mathcal{W} à partir de n et vice versa.

$$\text{card}(\mathcal{W}) = (n \times 2) + 1 \quad (3.2)$$

Le choix du paramètre n dépend du choix de description locale ou globale du contour. Plus n est grand, plus la description d'un point p_i se fait en fonction d'une plus grande portion du contour (cf. Figure 3.3). La valeur maximale prise par le paramètre n est $\frac{N-1}{2}$, et, dans ce cas, la description de chaque point du contour inclut tous les autres points sur le contour, il s'agit alors d'une description globale.

Chaque point est ensuite décrit par N_s triangles à différentes échelles. Dans la suite, N_s désigne d'une manière équivalente le nombre de triangles et le nombre d'échelles. Nous introduisons également une fonction de distance d_m où m est la distance entre deux triangles successifs. Cette fonction est soit linéaire d_m^{lin} soit logarithmique d_m^{log} . Ces deux types sont présentés dans les deux paragraphes suivants.

Fonction de distance linéaire

Étant donné un contour \mathcal{P} représenté par un ensemble de N points échantillonnés $\{p_1, \dots, p_N\}$. Une fonction de distance linéaire signifie l'utilisation de N_s triangles équidistants pour représenter chaque point du contour p_i . Formellement, elle est définie de la manière suivante :

$$d_m^{\text{lin}}(k) = k \times m \text{ avec } m \leq n \text{ et } k \geq 0 \quad (3.3)$$

Dans l'équation 3.3, le paramètre m définit la distance entre deux triangles successifs en terme de nombre de points $p \in \mathcal{P}$ (cf. Figure 3.4). m peut être considéré comme un paramètre de lissage. En effet, en prenant des triangles successifs proches, ce qui correspond à une faible valeur de m , nous favorisons la représentation des détails du contour. Par contre, plus la valeur de m est élevée, plus les triangles représentent une description grossière du contour.

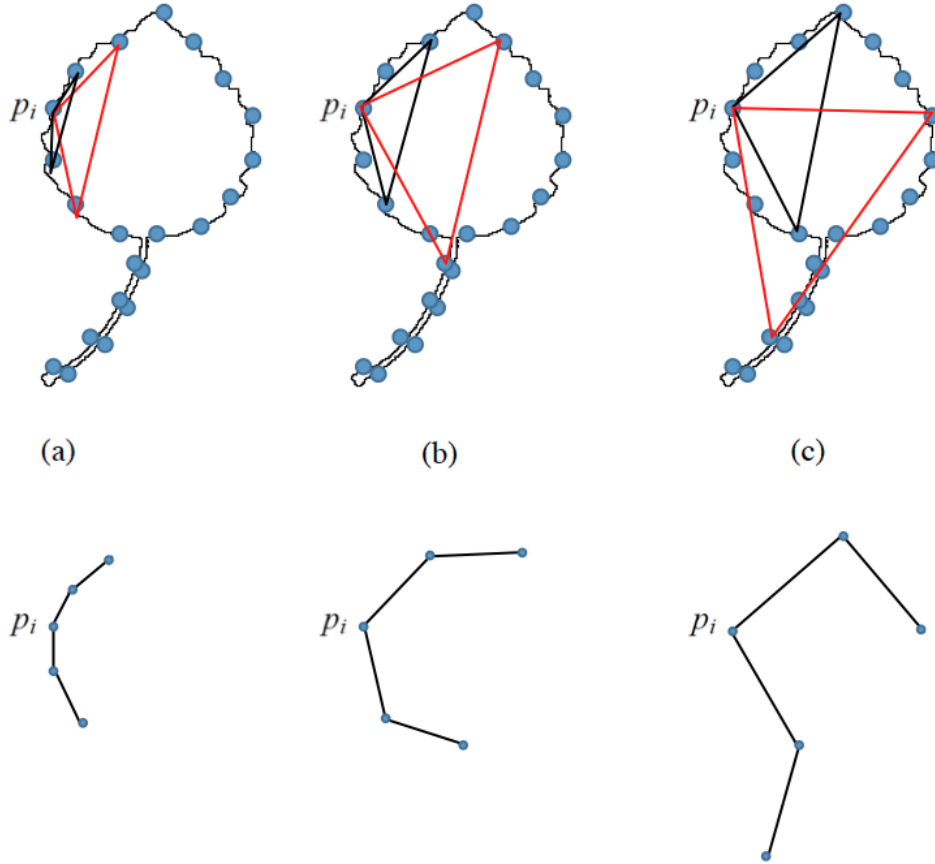


FIGURE 3.4 – Fonction de distance linéaire. Dans cet exemple $N_s = 2$. La première et la deuxième échelles sont représentées respectivement par le triangle noir et le triangle rouge (a) $d_1^{lin} : d(1) = 1, d(2) = 2$ (b) $d_2^{lin} : d(1) = 2, d(2) = 4$ (c) $d_3^{lin} : d_3^{lin}(1) = 3, d_3^{lin}(2) = 6$

fonction de distance logarithmique

Dans le cas de la distance logarithmique, nous utilisons une échelle logarithmique pour espacer les triangles successifs. L'idée sous-jacente est d'attribuer plus d'importance aux points voisins d'un point p_i qu'aux points plus éloignés. Elle est définie par l'équation suivante :

$$d_m^{log}(k) = m^k \text{ avec } m \leq n \quad (3.4)$$

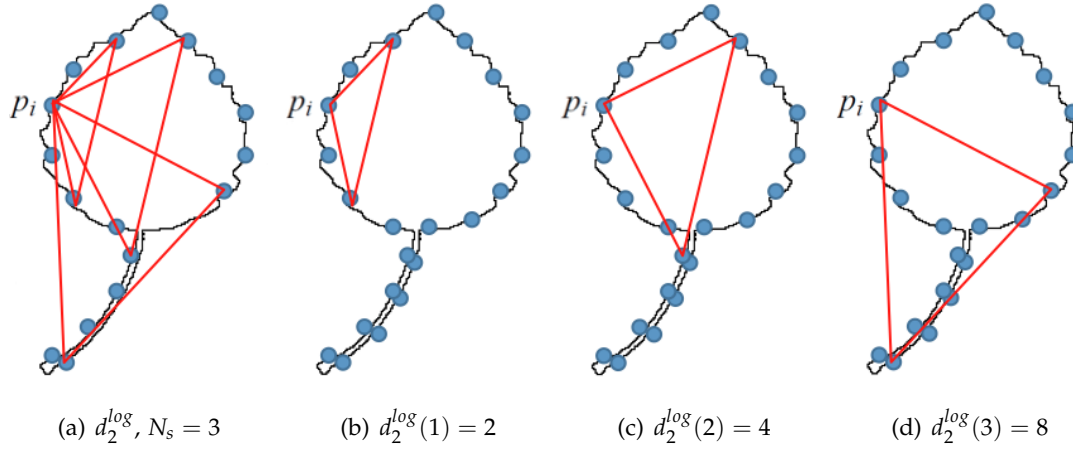


FIGURE 3.5 – Fonction de distance logarithmique.

Discussion des paramètres n, m, N_s

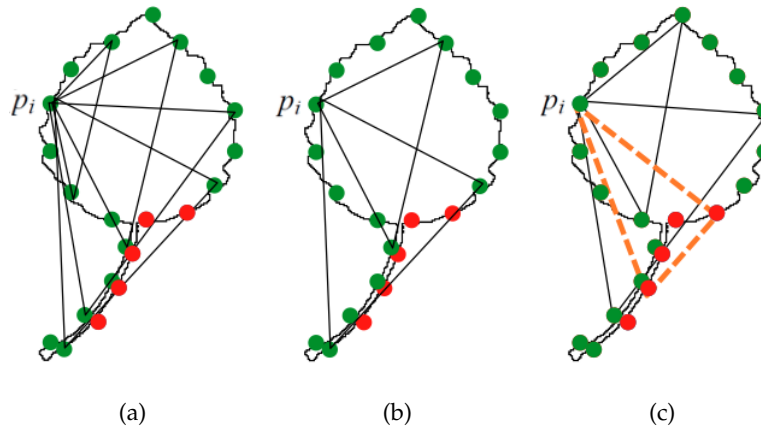
Les paramètres n, m et N_s sont auto-corrélés et chaque paramètre dépend des deux autres. En effet, si $m = 1$, cela signifie que $n = N_s$, c'est-à-dire, tous les triangles possibles dans le voisinage défini par n seront utilisés. Le nombre d'échelles N_s est donc toujours inférieur ou égal à n .

Dans notre implémentation, les deux paramètres m et n sont donnés en entrée de l'algorithme. Le nombre de triangles (d'échelles) utilisé peut être déduit à partir de ces deux paramètres. Par exemple :

- Si $n = 8$ et la distance entre deux triangles successifs est d_2^{lin} , le nombre d'échelles va être égal à 4 ($N_s = 4$) (voire figure 3.6(a)).
- Si $n = 8$ et la distance entre deux triangles successifs est d_4^{lin} , le nombre d'échelles va être égal à 2 ($N_s = 2$) (voire figure 3.6(b)).

En fait, nous définissons implicitement le nombre d'échelles N_s en spécifiant n et m . Généralement, n doit être divisible par m pour que les triangles puissent représenter l'intégralité de la portion du contour \mathcal{W} associée à n . Si ce n'est pas le cas, seulement les triangles formés par des points $p \in \mathcal{W}$ seront pris en compte. Par exemple :

- Si $n = 8$ et la distance entre deux triangles successifs est d_3^{lin} , le troisième triangle (triangle en orange sur dans figure 3.6(c)) n'est pas pris en compte car il est formé par des points n'appartenant pas tous à \mathcal{W} .

FIGURE 3.6 – Différentes valeurs de N_s en fonction de m et n

En résumé, la représentation triangulaire multi-échelle du contour, présentée dans ce paragraphe, comporte les paramètres suivants :

- N : le contour est initialement quantifié en N points
- n : définit le voisinage d'un point de contour p_i . Ce paramètre est exprimé en nombre de points de contour sélectionnés de chaque côté de p_i
- d_m : fonction de distance entre deux triangles successifs
- N_s : nombre de triangles (d'échelles) utilisé pour décrire chaque point du contour. Ce paramètre est déduit à partir de m et n .

Chaque point du contour p_i , est représenté par N_s triangles. Pour décrire le contour, il faut décrire ces triangles.

Dans le paragraphe 3.1.3, nous passons en revue le descripteur triangulaire de forme (TAR) puis nous introduisons de nouvelles descriptions triangulaires dans la section 3.1.4.

3.1.3 TAR (Triangle Area Representation)

Alajlan et al. [ARKF07] calculent, dans la description TAR, la surface signée de chaque triangle. Les deux avantages principaux de cette description sont :

- l'invariance aux transformations affines et la robustesse face au bruit et aux déformations du contour.
- le fait d'introduire le signe de la surface d'un triangle permet de détecter les lieux de concavité ou de convexité du contour. Une valeur négative, positive ou nulle de la surface du triangle impliquent respectivement l'existence d'une convexité, d'une concavité ou d'une ligne droite au voisinage de ce point.

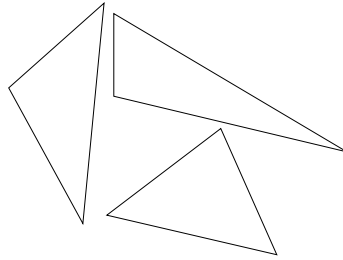


FIGURE 3.7 – Triangles de même surface

Supposons que $(x_{i-d(k)}, y_{i-d(k)})$, (x_i, y_i) et $(x_{i+d(k)}, y_{i+d(k)})$ soient les coordonnées respectives des points $p_{i-d(k)}$, p_i et $p_{i+d(k)}$. La surface signée du triangle T_i^k formé par ces trois points est donnée par la formule suivante :

$$\mathcal{A}(T_i^k) = \frac{1}{2} \begin{vmatrix} x_{i-d(k)} & y_{i-d(k)} & 1 \\ x_i & y_i & 1 \\ x_{i+d(k)} & y_{i+d(k)} & 1 \end{vmatrix} \quad (3.5)$$

Retenir la surface signée pour représenter un triangle permet de différencier les parties concaves ou convexes du contour mais ne retient aucune information sur la forme locale du contour. En effet, deux triangles de même surface peuvent avoir des formes différentes. C'est le cas des triangles de la figure 3.7. En fait, ces trois triangles ne sont pas semblables.

Deux triangles sont dits semblables¹ lorsqu'ils vérifient une des propriétés suivantes :

- (i) ils possèdent trois côtés dont les longueurs respectives sont proportionnelles,
- (ii) ils possèdent un angle de même amplitude compris entre deux côtés dont les longueurs respectives sont proportionnelles,
- (iii) ils possèdent deux angles de même amplitude deux à deux.

La donnée de la surface d'un triangle ne permet pas de connaître l'amplitude de ses angles ou les longueurs de ses côtés. Deux triangles de même surface ne vérifient *a priori* aucune des propriétés citées ci-dessus. Par conséquent, la description TAR n'est pas appropriée pour retrouver des triangles semblables. Dans ce qui suit, nous allons introduire trois nouvelles représentations triangulaires qui respectent les propriétés des triangles semblables.

1. cf. Groupe Cojerem, Géométrie en situations : 1er4e notions pour l'élève, De Boeck Wesmael edition, p. 58-60, 1995.

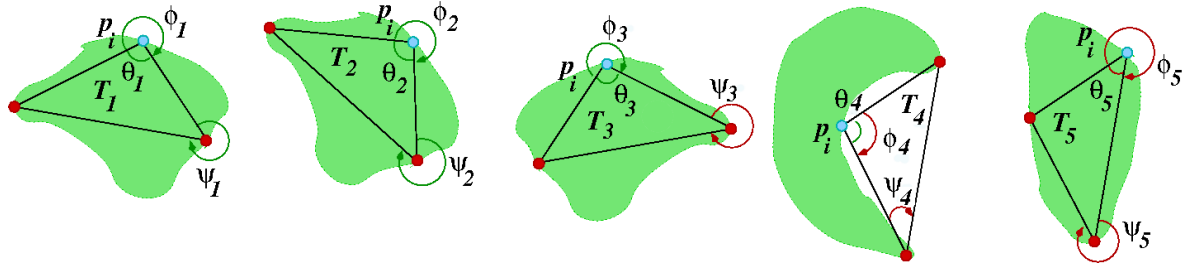


FIGURE 3.8 – Comparaison des représentations triangulaires

3.1.4 Descripteurs de forme proposés

TSL (Triangle Side Length Representation)

La représentation *TSL* est liée à la propriété (i) des triangles semblables. Elle retient les longueurs des côtés.

Soit le triangle T_i^k formé par les points du contour $p_{i-d(k)}$, p_i et $p_{i+d(k)}$, $k \in \{1, \dots, N_s\}$. Soient L_{1k} , L_{2k} et L_{3k} les longueurs des côtés du triangle T_i^k triés dans un ordre ascendant de longueur ($L_{1k} \leq L_{2k} \leq L_{3k}$). Nous définissons $M_k = L_{1k}/L_{3k}$ et $N_k = L_{2k}/L_{3k}$. Le descripteur *TSL* d'un triangle T_i^k est défini comme suit :

$$TSL(T_i^k) = (M_k, N_k) \quad (3.6)$$

Cette description est invariante à la translation et la normalisation par la longueur maximale des côtés assure l'invariance par rapport au changement d'échelle.

Examinons la figure 3.8 afin d'identifier les triangles qui ont la même description via *TSL*. Si nous considérons que T_1 est le triangle de base, T_2 est le résultat d'une rotation et T_3 est l'image miroir de T_1 . Ces deux dernières transformations préservent les distances entre les différents sommets du triangle T_1 . Le descripteur *TSL* est donc invariant à la rotation et à l'image miroir. Le triangle T_4 décrit une concavité du contour au point p_i à la différence de tous les autres triangles. La représentation *TSL* ne permet pas de repérer que le sommet p_i correspond à une concavité ou une convexité du contour. Le triangle T_5 a un angle au sommet p_i différent de tous les autres triangles mais il est représenté de la même façon que les autres parce que l'information de l'angle au sommet p_i est absente dans la description *TSL*. En fait, tous les triangles de la figure 3.8 sont semblables et ont la même description *TSL*.

TSLA (Triangle Side Length and an Angle Representation)

L'idée sous-jacente à cette description est de retenir le critère (ii) des triangles semblables en représentant un triangle par l'amplitude de l'angle interne au sommet p_i et les longueurs des deux côtés les plus petits du triangle. Nous partons de l'hypothèse que *TSLA* va enrichir la description *TSL* avec une information angulaire supplémentaire. En effet, l'angle au sommet p_i représente une information grossière sur la relation spatiale entre p_i et les deux autres sommets du triangle. L'ajout d'un angle pose plus de contraintes lors de la comparaison des formes et nous espérons que cela va apporter une certaine précision pour trouver des formes similaires.

Soit θ l'amplitude de l'angle interne au sommet p_i . La description *TSLA* peut alors s'écrire comme suit :

$$\text{TSLA}(T_i^k) = (M_k, N_k, \theta) \text{ avec } (M_k, N_k) = \text{TSL}(T_i^k) \quad (3.7)$$

De même que *TSL*, la description *TSLA* est invariante à la translation, à la rotation, au changement d'échelle et à l'effet miroir. Par contre, la signature *TSLA* n'est pas la même pour tous les triangles de la figure 3.8. Dans le cas du triangle T_5 , l'angle au sommet p_i est différent de tous les autres triangles, la signature *TSLA* de T_5 est donc différente de celle des autres.

TOA (Triangle Oriented Angles Representation)

Pour la description *TOA*, nous proposons d'utiliser les angles orientés du triangle. Soient $\phi_k = \widehat{p_{i-d(k)}p_i p_{i+d(k)}}$ et $\psi_k = \widehat{p_i p_{i+d(k)} p_{i-d(k)}}$ deux angles orientés consécutifs du triangle T_i^k . Le descripteur *TOA* d'un triangle T_i^k est défini par l'expression suivante :

$$\text{TOA}(T_i^k) = (\phi_k, \psi_k) \quad (3.8)$$

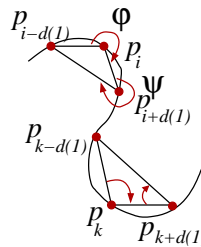


FIGURE 3.9 – Description TOA

Examinons la figure 3.8 pour décrire les propriétés du descripteur *TOA*. Les deux tri-

angles T_1 et T_2 ont la même signature TOA car la rotation préserve l’amplitude et l’orientation des angles. Contrairement à TSL et $TSLA$, l’encodage TOA du triangle T_3 est différent de celui de T_1 et T_2 . En effet, l’image miroir préserve les amplitudes d’angles mais pas leurs orientations. La signature TOA du triangle T_4 est également différente de celle des triangles T_1 et T_2 . Cela s’explique par le fait que nous calculons des angles orientés entre les vecteurs $\overrightarrow{p_{i-d(k)}p_i}$ et $\overrightarrow{p_ip_{i+d(k)}}$. L’orientation choisie est celle des aiguilles d’une montre. Cela va se traduire en pratique par des angles aigus ($\phi_k \leq 90^\circ, \psi_k \leq 90^\circ$) dans le cas où le triangle décrit une concavité locale et des angles obtus ($\phi_k \geq 90^\circ, \psi_k \geq 90^\circ$) dans le cas d’une convexité. Le descripteur TOA permet donc de distinguer les parties concaves des parties convexes du contour. La signature TOA de T_5 est différente de celle de T_1 et T_2 puisque les valeurs des angles aux sommets respectifs p_i sont différentes.

En conclusion, la description TOA est invariante à la translation, à la rotation et au changement d’échelle. Elle permet de coder différemment les concavités et les convexités du contour et elle est sensible à l’effet miroir.

Résumé des propriétés des descripteurs triangulaires proposées

Le tableau 3.1 résume les propriétés des différentes représentations triangulaires. Tous les descripteurs sont invariants par rapport à la translation, à la rotation et au changement d’échelle. La différence entre les descripteurs se situe au niveau de l’invariance par rapport à l’image miroir et la distinction entre les concavités et les convexités du contour.

Descripteur	invariance				concavités/convexités
	translation	rotation	échelle	image miroir	
TSL	✓	✓	✓	✓	
$TSLA$	✓	✓	✓	✓	
TOA	✓	✓	✓		✓

TABLE 3.1 – Propriétés des descriptions triangulaires : TSL , $TSLA$ et TOA

3.2 CONTEXTE DE FORME ÉTENDU ET POINTS SAILLANTS

La description de la forme foliaire peut être insuffisante dans le cas d’une forte similarité inter-espèce. Une information supplémentaire doit être incluse dans le processus d’identification. Dans cette section, nous présentons une approche de description des points saillants de la feuille. Les points saillants sont des points clefs de la feuille dont la position et les caractéristiques visuelles locales peuvent s’avérer utiles pour l’identification d’espèces vé-

gétales.

L'approche d'identification basée sur les points saillants, que nous proposons, comporte deux étapes :

- Détection de points clefs. Dans notre cas, un détecteur générique de points d'intérêt est utilisé, à savoir, le détecteur de Harris.
- Description des points saillants. Nous avons choisi de considérer deux façons de décrire les points extraits : leur répartition spatiale via un contexte de forme (Shape Context) étendu. Nous enrichissons ensuite cette description par le calcul des descripteurs locaux dans un voisinage autour des points d'intérêt. Cela vise à décrire la micro-texture et la forme locale à l'intérieur des petites fenêtres (patches) ayant pour origine les points saillants détectés.

3.2.1 Points saillants d'une feuille en botanique

En botanique, un point saillant d'une feuille est un point particulier présent sur la grande majorité des feuilles et facilement identifiable par les experts. Deux points de la feuille sont communément identifiés par les botanistes : la *base* qui est le point d'attache entre le pétiole et le limbe et l'*apex* qui représente le sommet du limbe. En effet, les botanistes s'intéressent particulièrement à la forme du contour au voisinage de ces deux points. Ceci est illustré dans la figure 3.10.



FIGURE 3.10 – Régions basales (première ligne) : cunéiforme, convexe et arrondie et régions à l'apex (deuxième ligne) : droite, convexe et arrondie

La base et l'apex ne sont pas les points uniques utilisés par les botanistes. Ceux-ci ont généralement recours à un ensemble plus large de points saillants pour identifier l'espèce d'une feuille. Cet ensemble inclut les points spéciaux sur le contour (sommets des lobes, points de changement de courbure, etc.), les points de jonction et de terminaison des nervures principales ou secondaires, etc. (cf. Figure 3.11).

La définition botanique des points saillants peut être transposée dans le domaine de

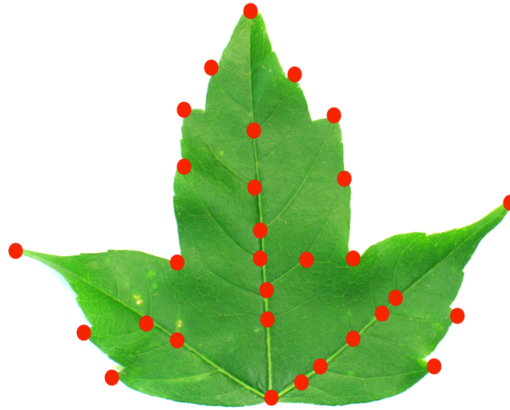


FIGURE 3.11 – Exemple de points saillants d’une feuille

la vision par ordinateur. L’une des techniques les plus répandues est celle basée sur les points d’intérêt. Les points d’intérêt sont situés à des endroits caractéristiques dans une image comme par exemple les coins, les points d’intersection, etc. La localisation de ce type de points est tout à fait pertinente pour décrire les images de feuilles. En effet, de nombreuses feuilles se caractérisent par des variations importantes sur leurs contours (feuilles dentées ou lobées) [MA04] ou par des points de jonction sur leur nervation [PHN06, PHN08, NHK08].

Nous présentons dans le paragraphe suivant, la méthode de détection de points saillants que nous avons utilisée.

3.2.2 Détection des points saillants

Les points saillants de la feuille proviennent des caractères foliaires différents (contour, nervation, texture, etc.) et ne possèdent pas forcément les mêmes caractéristiques visuelles. Il est donc très difficile de concevoir une méthode de détection de tous ces points. Nous avons donc choisi de les approximer par un détecteur générique de points d’intérêt et plus particulièrement d’utiliser un détecteur de coins.

Ce choix est motivé par le fait que les coins sont assez représentatifs des variations du contour. De plus, ils sont appropriés pour trouver les jonctions du réseau de nervures.

L’un des premiers détecteurs de points d’intérêt est celui Moravec [Mor77]. Il s’agit d’un détecteur de coins fonctionnant sur des images prises par une caméra implantée sur un robot, dans le but de le guider de façon à éviter les obstacles. Un grand nombre de travaux

proposant des améliorations du détecteur de Moravec a été proposé [Bea78, KR82, HS88]. Dans le paragraphe suivant, nous nous focalisons sur la présentation du détecteur de Harris que nous avons retenu.

Détecteur de Harris

Harris et Stephen ont corrigé certaines limitations du détecteur de Moravec pour proposer le détecteur de coins populaire : le détecteur de Harris.

Une étude comparative de plusieurs détecteurs de points d'intérêt a été menée par Schmid et al [SMB98, SMB00]. Les auteurs ont conclu que le détecteur de Harris est l'un des plus stables en terme de répétabilité. Ce critère de base pour l'évaluation des détecteurs de points d'intérêt consiste à vérifier que le même ensemble de points est détecté sous certaines transformations de l'image : rotation, changement de luminance, etc.

Le critère de répétabilité est important dans la mesure où les bases d'images de feuilles considérées contiennent généralement des feuilles avec rotation et changement d'échelle.

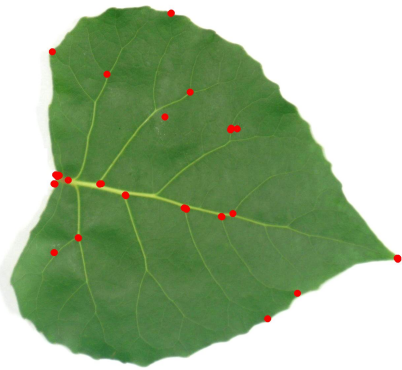
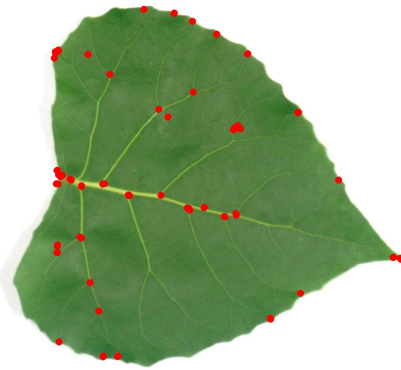
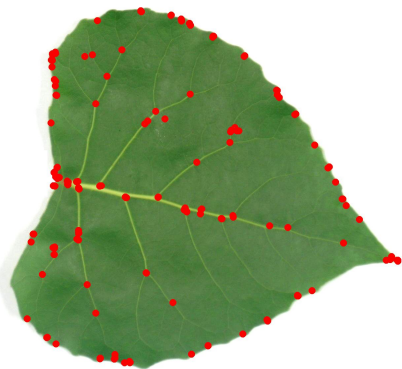
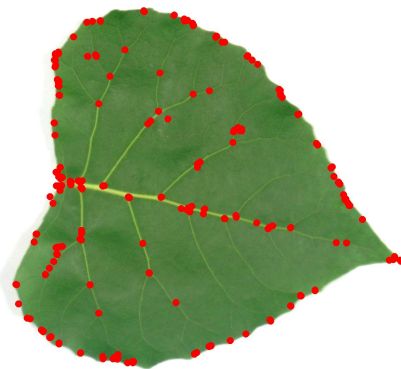
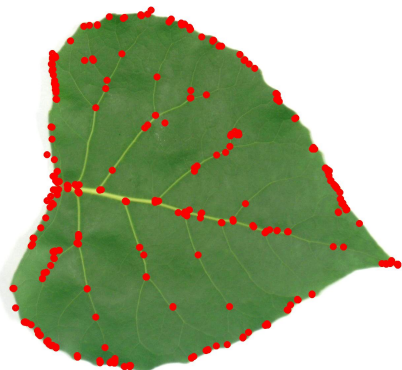
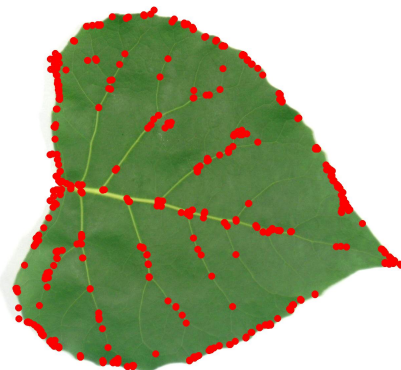
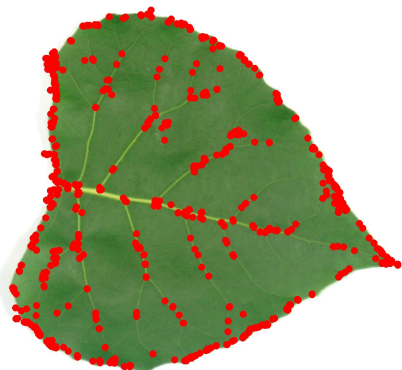
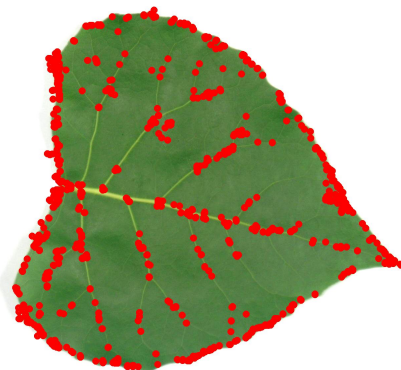
Nous avons utilisé la version du détecteur de Harris implémentée dans IKONA/maestro [BFF⁺01] qui possède les caractéristiques suivantes :

- un détecteur de Harris sur les images en couleur proposé par Valérie Gouet dans [Gou00].
- un détecteur de Harris multi-échelle.

Notre choix a été conforté par les évaluations menées au sein de l'équipe IMEDIA dans le cadre de la campagne d'évaluation ImageCLEF [GBJ⁺11]. Le but était d'étudier la pertinence des détecteurs de points d'intérêt Harris et SIFT/SURF dans le contexte particulier d'identification d'espèces. Les évaluations ont montré que les descripteurs calculés sur les points de Harris obtiennent le meilleur score d'identification.

Le nombre de points saillants extraits, que nous notons N_{ps} , représente un paramètre important. La figure 3.12 montre les points saillants détectés en faisant varier N_{ps} . Nous remarquons que lorsque le nombre N_{ps} est très élevé ($N_{ps} > 600$), beaucoup de points "bruit" sont détectés, c'est à dire, des points qui ne sont pas forcément caractéristiques de la feuille.

Dans le cas inverse où le nombre de points de Harris est réduit ($N_{ps} < 200$), cela approxime assez bien une grande partie des points clefs de la feuille. Cette observation est générale mais dépend étroitement du type et de la morphologie de la feuille.

(a) $N_{ps} = 50$ (b) $N_{ps} = 100$ (c) $N_{ps} = 200$ (d) $N_{ps} = 300$ (e) $N_{ps} = 400$ (f) $N_{ps} = 600$ (g) $N_{ps} = 800$ (h) $N_{ps} = 1000$ FIGURE 3.12 – Variation du nombre de points saillants N_{ps} détectés.

En effet, nous avons observé que les points de Harris sont situés essentiellement sur le contour si celui-ci présente de fortes variations (cas des feuilles dentées). Ceci est illustré dans la figure 3.13. Dans le cas où le contour est lisse ou présente de faibles variations, les points de Harris sont répartis entre le contour et l'intérieur du limbe. Les points internes au limbe sont situés généralement sur la nervation ou les endroits de micro-texture (cf. Figure 3.14).

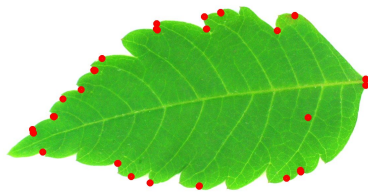
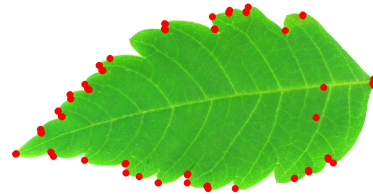
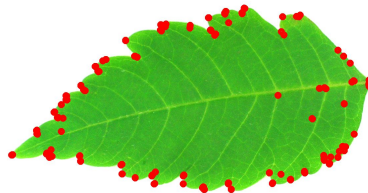
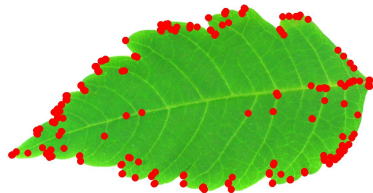
(a) $N_{ps} = 50$ (b) $N_{ps} = 100$ (c) $N_{ps} = 200$ (d) $N_{ps} = 300$

FIGURE 3.13 – Points de Harris détectés sur une feuille dentée.

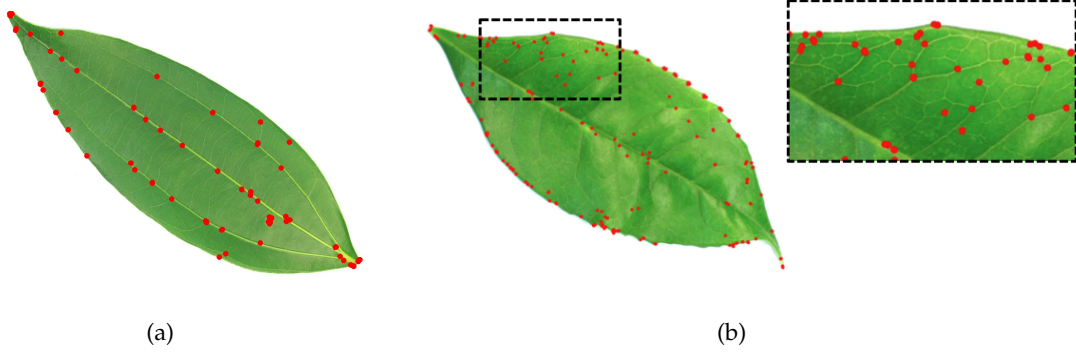


FIGURE 3.14 – Points de Harris sur la nervation et sur les éléments de texture. (a) points positionnés sur la nervation (b) points positionnés sur la micro-texture (micro-nervures).

Après avoir détecté les points saillants avec le détecteur de Harris, il convient de les décrire. Ceci est détaillé dans le paragraphe suivant.

3.2.3 Description des points saillants

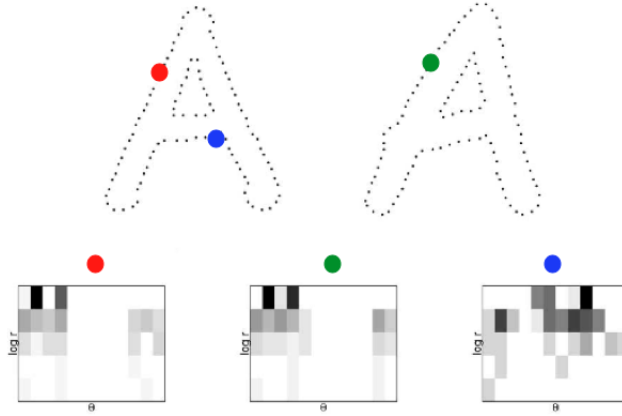
Pour décrire l'ensemble de points saillants extraits, nous proposons un descripteur basé sur le contexte de forme de Belongie et al. [BMP02] (cf. Section 3.2.3).

Contexte de forme (*Shape Context*)

Le descripteur de forme *Shape Context* (ou contexte de forme) a été proposé par Belongie et al. [BMP02]. L'idée de ce descripteur est de considérer un ensemble de N points uniformément distribués sur le contour $\mathcal{C} = \{p_1, \dots, p_i, \dots, p_N\}$ et de représenter chaque point p_i par la distribution spatiale des autres points vus de ce point. La distance euclidienne et l'angle sont calculés entre chaque paire de points du contour. Les distances et les directions sont quantifiées respectivement en N_D et N_A intervalles dans un repère log-polaire. Cette méthode de quantification est utilisée par certains travaux [ASR07, LJ07, SI07]. Le descripteur *Shape Context* est stocké dans un vecteur 2D de dimension $N_D \times N_A$ où chaque case contient le nombre de points du contour appartenant au *bin* correspondant dans la représentation log-polaire. Ainsi, deux points ayant une configuration spatiale proche auront des histogrammes similaires. (cf. Figure 3.15).

Contexte de forme étendu (*Advanced Shape Context ASC*)

Dans la description shape context, le même ensemble de points \mathcal{C} est utilisé à la fois pour :

FIGURE 3.15 – La description *Shape Context*. Figure dans [BMP02]

- calculer les histogrammes spatiaux pour chaque $p_i \in \mathcal{C}$.
- voter dans la description spatiale de chaque point p_i , c'est à dire, le calcul d'un histogramme shape context de p_i utilise des points p_j de $\mathcal{C} : \{p_j \in \mathcal{C} \mid p_j \neq p_i\}$.

Dans ce paragraphe, nous proposons un descripteur *ASC* (Advanced Shape Context) qui étend le contexte de forme classique en distinguant deux ensembles :

- un ensemble de calcul \mathcal{C} : un histogramme est attribué à chaque point $p_i \in \mathcal{C}$.
- un ensemble votant \mathcal{V} . Le calcul de l'histogramme de chaque point $p_i \in \mathcal{C}$ se fait en fonction des points $p_j \in \mathcal{V}$

Dans le schéma *ASC*, et contrairement au descripteur Shape Context initial, les ensembles \mathcal{C} et \mathcal{V} peuvent être distincts.

Étant donné un ensemble de n points \mathcal{V} et un point p de \mathbb{R}^2 ($p \in \mathcal{C}$), le descripteur *ASC* de \mathcal{C} au point p est une représentation discrète de l'ensemble des n vecteurs définis par les paires de points (p, q) avec $q \in \mathcal{V}$. $ASC(p, \mathcal{V})$ est encodé par un histogramme où chaque paire de points (p, q) représentée par une distance r et un angle θ , contribue dans le bin k défini en utilisant la quantification log-polaire introduite dans [BMP02]. Formellement, le descripteur *ASC* d'un point p associé à un ensemble \mathcal{V} est donné par l'équation suivante :

$$ASC(p, \mathcal{V})_k = \{q \in \mathcal{V} : q - p \in bin_p(k)\} \quad (3.9)$$

Dans la suite, nous adoptons la notation suivante :

$$ASC_{\mathcal{C}-\mathcal{V}}$$

pour désigner le descripteur de contexte de forme étendu où \mathcal{C} est l'ensemble de calcul et \mathcal{V} est l'ensemble votant. En faisant varier les points de \mathcal{C} et de \mathcal{V} , nous définissons les trois scénarios suivants :

ASC_{PC-PC} **Relations spatiales entre les points du contour.**

Dans ce scénario, l'ensemble de calcul \mathcal{C} et l'ensemble votant \mathcal{V} sont identiques et sont référés par $PC = \{\text{points quantifiés sur le contour}\}$. Chaque ensemble contient les points uniformément quantifiés sur le contour.

$$\mathcal{C} = \mathcal{V} = PC = \{\text{points échantillonnés sur le contour}\}$$

C'est une description identique à celle du contexte de forme proposée par Belongie et al. [BMP02]. Les processus complets (description + mise en correspondance) sont différents puisque Belongie et al. utilisent une approche de comparaison globale des formes alors que nous allons utiliser une méthode de mise en correspondance locale qui cherche les points du contour ayant des signatures similaires.

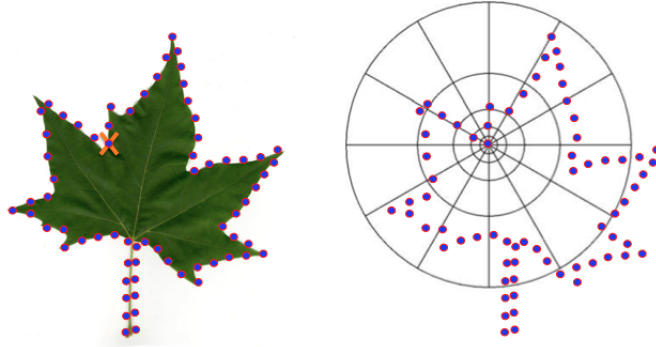


FIGURE 3.16 – Scénario ASC_{PC-PC}

ASC_{PS-PS} **Relations spatiales entre les points saillants.**

Comme dans le scénario précédent, \mathcal{C} et \mathcal{V} contiennent le même ensemble de points, constitué des points saillants trouvés par le détecteur de Harris.

$$\mathcal{C} = \mathcal{V} = PS = \{\text{points saillants détectés avec le détecteur de Harris}\}$$

D'une façon similaire à l'approche de recherche de logos [RnL10], nous cherchons, à travers ce scénario, à étudier le rôle et l'impact des relations spatiales entre les points saillants dans l'identification d'espèces végétales. En particulier, nous étudions

la pertinence de la caractérisation d'une feuille par les relations spatiales entre ses points saillants.

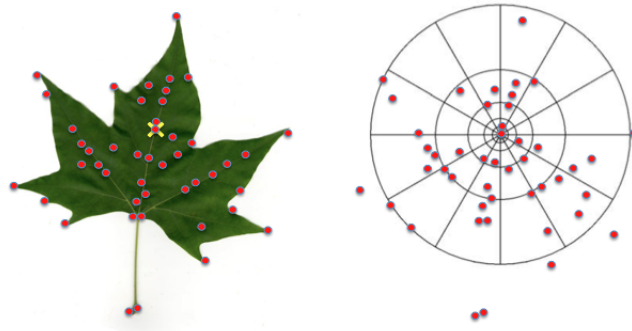


FIGURE 3.17 – Scénario ASC_{PS-PS}

ASC_{PS-PC} Corrélation spatiale entre les points saillants et le contour

En botanique, il existe un lien sémantique entre la nervation et la forme. Nous cherchons à travers ce scénario à exploiter cette information en représentant les points saillants d'une feuille dans un contexte défini par son contour. Ce scénario vise à étudier la corrélation spatiale entre les points d'intérêt et les points du contour de la feuille.

L'ensemble de calcul est formé par les points de Harris et l'ensemble votant \mathcal{V} est composé de points du contour. Contrairement au scénario ASC_{PC-PC} , où l'ensemble PC contient les points quantifiés sur le contour, PC est composé ici de tous les points du contour.

$$\mathcal{C} \neq \mathcal{V}, \mathcal{C} = PS = \{\text{Points saillants}\} \text{ et } \mathcal{V} = PC = \{\text{Tous les points du contour}\}$$

Descripteurs locaux

Le descripteur ASC_{PS-PC} représente une information spatiale impliquant le contour et les points d'intérêt mais ne contient aucune information sur la micro-texture locale dans le voisinage des points saillants. Afin d'enrichir cette description, un ensemble de descripteurs locaux est calculé sur un petit voisinage centré en chaque point saillant. Les descripteurs locaux utilisés sont :

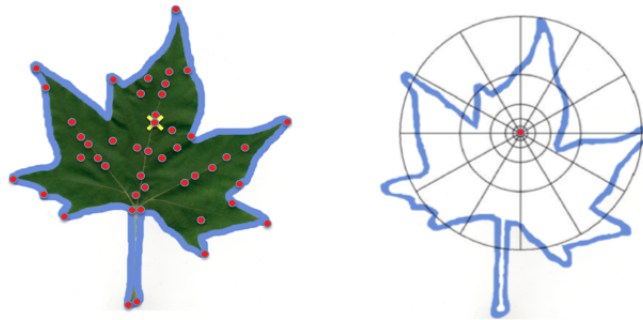


FIGURE 3.18 – Scénario ASC_{PS-PC} . Les points rouges constituent l'ensemble de clacul \mathcal{C} et les points en bleu constituent l'ensemble votant \mathcal{V}

- un histogramme de Hough de taille 16 basé sur la transformée de Hough introduite par Paul Hough en 1962. Elle est utilisée pour détecter des formes élémentaires dans une image [Fer05] tels qu'une droite, cercle, ellipse, etc.
- un histogramme de Fourier basé sur la transformée de Fourier [Fer05] qui encode une information relative à la fréquence des motifs élémentaires.
- un histogramme des directions du gradient

Afin d'assurer l'invariance à la rotation, le voisinage des points saillants est représenté par des petites fenêtres orientées selon la direction principale du gradient à l'intérieur de ces fenêtres. Ces trois descripteurs locaux associés à des points de Harris ont obtenu un bon score d'identification sur les scans de feuilles dans [GJY⁺11].

Un quatrième scénario est proposé, il consiste à combiner le scénario ASC_{PS-PC} et les descripteur locaux, cités ci-dessus, autour des points saillants. La combinaison revient à concaténer les vecteurs associés. Ce scénario est noté $eASC_{PS-PC}$ (enriched ASC_{PS-PC}).

Le tableau 3.2 résume les quatre scénarios proposés.

Scénario	Ensemble de calcul \mathcal{C}	Ensemble votant \mathcal{V}	But
ASC_{PC-PC}	$\mathcal{C} = \{\text{points du contour}\}$	$\mathcal{V} = \{\text{points du contour}\}$	évaluation de la description Shape Context
ASC_{PS-PS}	$\mathcal{C} = \{\text{points saillants}\}$	$\mathcal{V} = \{\text{points saillants}\}$	étude de la pertinence de la relation spatiale entre les points saillants
ASC_{PS-PC}	$\mathcal{C} = \{\text{points saillants}\}$	$\mathcal{V} = \{\text{points du contour}\}$	étude de la corrélation spatiale entre les points saillants d'une feuille et son contour
$eASC_{PS-PC}$	$\mathcal{C} = \{\text{points saillants}\}$	$\mathcal{V} = \{\text{points du contour}\}$	enrichissement de la description ASC_{PS-PC} par des descripteurs locaux autour des points saillants pour décrire la micro-texture

TABLE 3.2 – Scénarios étudiés

3.3 EXTRACTION ET DESCRIPTION DE LA NERVATION

Les nervures constituent un caractère important de la feuille. Leur extraction représente un véritable défi car leur apparence visuelle peut varier d'une image à une autre. De plus, la netteté du réseau de nervures dépend à la fois du protocole d'acquisition des images et des facteurs naturels (âge de la feuille, feuille morte, etc.). Ceci est illustré dans la figure 3.19.



FIGURE 3.19 – Exemples d'images de feuilles présentant une difficulté d'extraction de nervures.

Dans la littérature la nervation a rarement été utilisée comme caractère principal pour l'identification des espèces végétales (voir chapitre 2). Elle est souvent considérée comme une source d'information supplémentaire et, comme la couleur, elle est utilisée en combinaison avec des caractères plus discriminants de la feuille [KNSS11b, NHK08, PHN08]. Dans ce chapitre, nous proposons une approche d'identification basée principalement sur la

description du réseau de nervures. Nous présentons une méthode d'extraction de nervures où des outils de la morphologie mathématique sont utilisés, et une méthode de description qui considère plusieurs aspects de la nervation d'une feuille : les principales directions du réseau de nervures, son arrangement spatial et sa corrélation spatiale dans un contexte défini par le contour.

3.3.1 Opérateurs morphologiques

Les nervures d'une feuille peuvent être vues en vision par ordinateur comme des structures linéaires contrastées. Pour pouvoir les extraire, il est nécessaire d'utiliser les outils appropriés. Nous avons choisi une approche morphologique pour l'extraction des nervures en raison de la richesse des opérateurs qu'elle offre. En effet, les opérateurs morphologiques peuvent jouer le rôle d'un filtre afin d'améliorer la qualité des images mais ils permettent aussi d'extraire des éléments d'une certaine structure (cf. Paragraphe 3.3.2) grâce à la notion d'élément structurant que nous expliquons dans la suite.

Un opérateur morphologique utilise un élément structurant que nous allons noter B . Celui-ci est défini par un voisinage de pixels qui a pour origine un pixel de référence. B est caractérisé par sa forme et sa taille. Le choix de la forme et de la taille de l'élément structurant est important et dépend de l'application. La figure 3.20 montre trois exemples d'éléments structurants de forme différente. Tous les opérateurs morphologiques sont fondés sur la

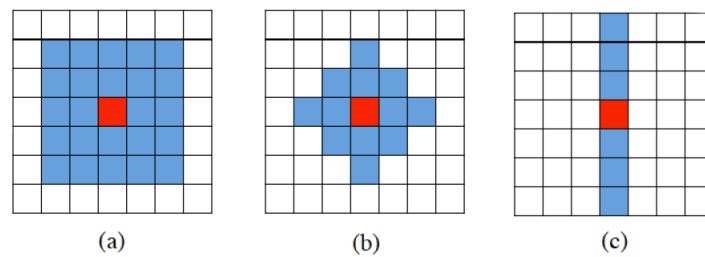


FIGURE 3.20 – Éléments structurants. Pixel de référence en rouge et voisinage de pixels en bleu. (a) : élément structurant carré de taille 5. (b) : élément structurant hexagonal de taille 5. (c) : élément structurant linéaire de taille 7.

sélection d'un élément structurant.

Les deux opérateurs morphologiques de base sont l'érosion et la dilatation. Un grand nombre d'opérateurs avancés peuvent être définis à partir de ceux-ci. Nous présentons dans la section suivante les principaux opérateurs utilisés.

Érosion

L'érosion ensembliste consiste à poser en chaque pixel x d'un objet X la question : *l'élément structurant B_x est-il contenu entièrement dans X ?*

L'ensemble érodé est constitué des points où la réponse à cette question est affirmative. L'ensemble initial représente souvent les objets d'une image binaire ou le sous graphe d'une image en niveaux de gris.

L'érosion d'un ensemble X par un élément structurant B , notée $\epsilon_B(X)$, est définie par l'ensemble des points, x , tels que B est inclus dans X quand son origine est placée en x .

$$\epsilon_B(X) = \{x \mid B_x \subset X\} \quad (3.10)$$

L'équation 3.10 peut aussi être écrite sous forme d'intersection d'un ensemble de translations. Ces translations sont définies par l'élément structurant.

$$\epsilon_B(X) = \bigcap_{b \in B} X_{-b} \quad (3.11)$$

L'érosion peut être appliquée à des images binaires ou à des images en niveaux de gris : l'érosion d'une image I par un élément structurant B , notée $\epsilon_B(I)$, est définie comme étant le minimum des translations de I par les vecteurs $-b$ de B .

$$\epsilon_B(I) = \bigwedge_{b \in B} I_{-b} \quad (3.12)$$

La valeur érodée d'un pixel donné x est la valeur minimum de l'image dans la fenêtre définie par l'élément structurant quand son origine est placée en x :

$$\epsilon_B(I)(x) = \min_{b \in B} I(x + b) \quad (3.13)$$

L'érosion réduit les « pics » de niveaux de gris et élargit les « vallées » : elle tend donc à homogénéiser l'image, à l'assombrir et à étaler le bord des objets les plus sombres comme le montre la figure 3.21

Dilatation

La dilatation est l'opérateur dual de l'érosion. Elle consiste à poser en chaque pixel x d'un objet X , la question : *"l'élément structurant B_x intersecte-t-il l'ensemble X ?"*

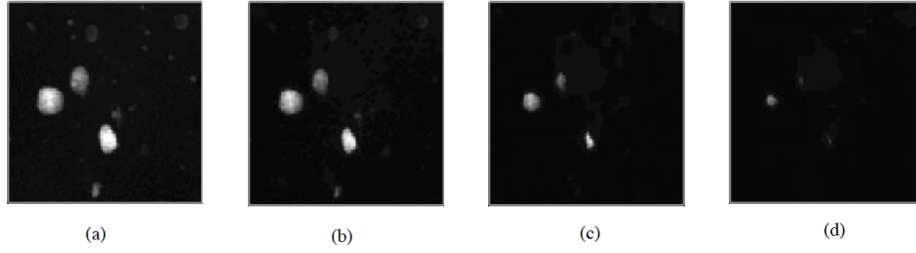


FIGURE 3.21 – Érosion morphologique. (a) image originale en niveaux de gris ;(b) érosion avec un élément structurant carré de taille 3 ; (c) érosion avec un élément structurant carré de taille 6 ; (d) érosion avec un élément structurant carré de taille 10.

L'ensemble dilaté est constitué des points où la réponse à cette question est l'affirmative. La dilatation d'un ensemble X par un élément structurant B , notée $\delta_B(X)$, est définie par l'ensemble de points x tels que B intersecte X quand son origine coïncide avec x :

$$\delta_B(X) = \{x \mid \check{B}_x \cap X \neq \emptyset\} \quad (3.14)$$

Dans l'équation 3.14, \check{B} désigne l'élément structurant symétrique de B . Notons que la dilatation d'un seul point donne comme sortie l'élément structurant centré sur son origine. L'équation 3.14 peut aussi s'écrire sous forme d'unions de translations. Ces translations sont définies par l'élément structurant.

$$\delta_B(X) = \bigcup_{b \in B} X_b \quad (3.15)$$

La définition précédente peut être appliquée sur des images binaires ainsi que sur des images en niveaux de gris : la dilatation d'une image I par un élément structurant B , notée $\delta_B(I)$, est définie par le maximum de la translation de I par les vecteurs b de B .

$$\delta_B(I) = \bigvee_{b \in B} I_b \quad (3.16)$$

La valeur dilatée d'un pixel x donné est le maximum de l'image dans la fenêtre définie par l'élément structurant quand son origine est placée en x :

$$[\delta_B(I)](x) = \max_{b \in B} I(x + b) \quad (3.17)$$

Cette transformation comble les « vallées » et épaissit les « pics » : éclaircit l'image et tend à faire disparaître les objets sombres.

La dilatation et l'érosion sont des transformations adjointes et respectent le principe

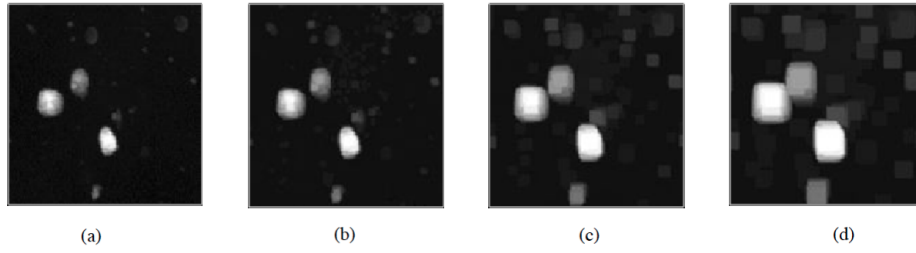


FIGURE 3.22 – Dilatation morphologique. (a) image originale en niveaux de gris ; (b) dilatation avec un élément structurant carré de taille 3 ; (c) dilatation avec un élément structurant carré de taille 6 ; (d) dilatation avec un élément structurant carré de taille 10.

de la complémentarité. Cela signifie qu'une érosion sur une image est équivalente à la dilatation sur l'image inverse avec le même élément structurant (et vice versa). Ces deux transformations permettent d'introduire les opérateurs d'ouverture et de fermeture morphologique.

Ouverture

L'ouverture d'une image I par un élément structurant B , notée $\gamma_B(I)$, est définie par l'érosion de I par B suivie d'une dilatation par l'élément structurant symétrique \check{B} .

$$\gamma_B(I) = \delta_{\check{B}} \circ \epsilon_B(I) \quad (3.18)$$

L'ouverture supprime les pics mais préserve les vallées. Elle homogénéise l'image mais préserve les objets sombres.

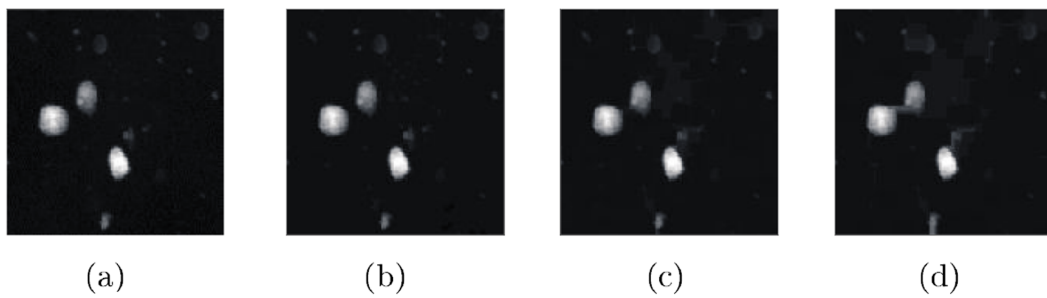


FIGURE 3.23 – Ouverture morphologique. (a) image en niveaux de gris originale ; (b) ouverture avec un élément structurant carré de taille 3 ; (c) ouverture avec un élément structurant carré de taille 6 ; (d) ouverture avec un élément structurant carré de taille 10.

Fermeture

La fermeture d'une image I par un élément structurant B , notée $\varphi_B(I)$, est définie par la dilatation de I par un élément structurant B , suivie de l'érosion par l'élément structurant symétrique \check{B} :

$$\varphi_B(I) = \epsilon_{\check{B}} \circ \delta_B(I) \quad (3.19)$$

La fermeture comble les vallées. Elle homogénéise et éclaircit l'image, comme le montre la figure 3.24.

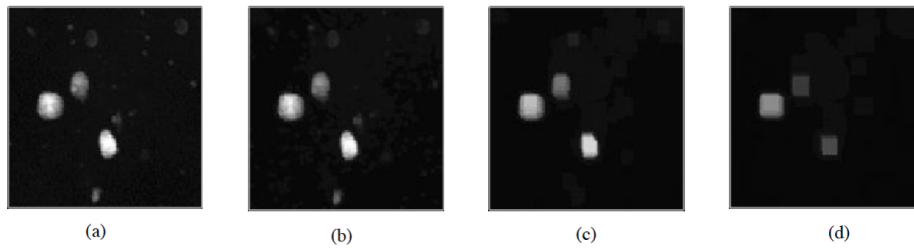


FIGURE 3.24 – Fermeture morphologique. (a) image en niveaux de gris originale ; (b) fermeture avec un élément structurant carré de taille 3 ; (c) fermeture avec un élément structurant carré de taille 6 ; (d) fermeture avec un élément structurant carré de taille 10.

3.3.2 Algorithme d'extraction de nervures

L'algorithme d'extraction de nervures comprend les étapes suivantes :

- **Pré-filtrage.** Cette étape vise à améliorer le contraste dans les images, en particulier, le contraste des nervures par rapport à la couleur globale du limbe.
- **Extraction des structures contrastées.** Nous gardons dans cette étape l'ensemble des structures présentant un contraste local. Le degré du contraste n'est pas mesuré : nous traitons d'une façon équivalente les structures de faible ou de fort contraste. En effet, il est très rare que la totalité du réseau de nervures soit fortement contrasté. Pour cette raison, nous nous intéressons également aux structures de faible contraste afin d'extraire un maximum d'information.
- **Détection des structures linéaires.** Nous nous intéressons dans cette étape aux petits éléments contrastés dont la forme est linéaire. Pour ce faire, nous partons de l'hypothèse que le réseau de nervures peut être approximé par un ensemble de segments. Le résultat de cette étape est une première détection des nervures avec un nombre important de faux positifs.

– **Post-filtrage.** Celui-ci vise à éliminer les structures n'appartenant pas aux nervures. Le processus décrit ci-dessus est représenté dans la Figure 3.25.

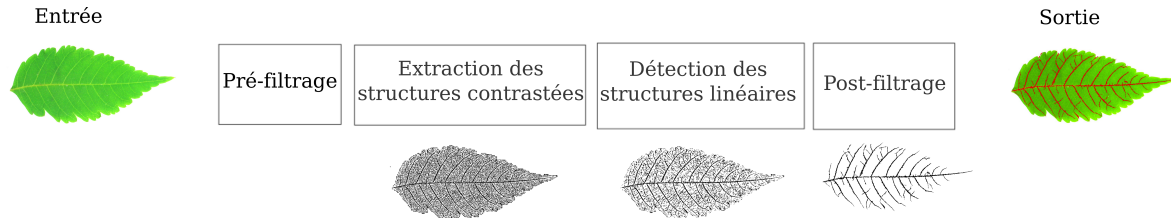


FIGURE 3.25 – Processus d'extraction des nervures

Les trois premières étapes de l'algorithme sont similaires à celles présentées dans [EM12]. Les travaux effectués dans [EM12] entrent dans le cadre d'un stage précédent où nous avons cherché à extraire les éléments saillants des feuilles de vigne (nervures, base, lobes, etc.). L'idée de détecter des structures linéaires à partir des structures contrastées est similaire à celle dans [EM12]. Cependant, l'algorithme global d'extraction est différent. Les principales différences sont les suivantes :

- Les structures linéaires détectées dans [EM12] sont représentées à l'aide des polynômes de deuxième degré en utilisant une régression linéaire. Dans notre algorithme, et pour des raisons de simplicité, les structures linéaires sont considérées comme de simples composantes connexes.
- Un seuil global a été utilisé dans [EM12] pour extraire les structures linéaires au-dessus d'une certaine taille. Cela a généré un réseau de nervures déconnecté qui a nécessité une phase de reconstruction. Cette démarche est très coûteuse en temps de calcul. L'algorithme d'extraction que nous introduisons ici propose une nouvelle méthode de filtrage basée sur le calcul d'un seuil adaptatif. Nous avons évité ainsi l'étape de reconstruction des nervures déconnectées qui, en plus de son coût considérable, n'est pas toujours efficace.
- L'objectif étant différent : dans [EM12], le but était d'identifier les nervures principales parmi l'ensemble de nervures extraites, afin de détecter les sommets des lobes. Notre objectif ici est d'utiliser l'intégralité du réseau de nervures pour étudier son impact dans l'identification d'espèces.
- Finalement, notre travail précédent était spécifique aux feuilles ayant une nervation palmée alors que nous cherchons ici à développer un algorithme d'extraction générique à tous les types de nervation.

Dans les prochains paragraphes, nous détaillons la description de chaque étape de l'algorithme d'extraction de nervures.

Amélioration de la qualité d'images

Une première étape de pré-filtrage est nécessaire pour améliorer la qualité d'images. Pour ce faire, Meyer et Serra [MS89] ont introduit l'opérateur morphologique *toggle-mappings*. Celui-ci vise à augmenter la netteté des structures contenues dans l'image. Nous travaillons sur l'image d'intensité I en sélectionnant le canal vert du système de couleur RGB. Le toggle-mappings est appliqué à chaque pixel p de I à partir d'une transformée extensive T_1 et une transformée anti-extensive T_2 de la façon suivante :

$$I_{TM} = [TM(I)](p) = \begin{cases} [T_1(I)](p) & \text{si } [T_1(I) - I](p) \leq [I - T_2(I)](p) \\ [T_2(I)](p) & \text{sinon.} \end{cases} \quad (3.20)$$

Cet opérateur remplace la valeur de chaque pixel p d'une image I par la valeur prise parmi $[T_1(I)](p)$ ou $[T_2(I)](p)$ qui lui est la plus proche. Cela a pour effet de mieux mettre en valeur le changement d'intensité dans un voisinage de pixels et par conséquent d'augmenter la netteté des petites structures.

Dans notre cas, d'une manière similaire à Meyer et Serra [MS89] et Walter [Wal03], une ouverture est utilisée comme transformée extensive $T_1 = \gamma$ et une fermeture est utilisée comme transformée anti-extensive $T_2 = \varphi$.

L'étape suivante, décrite ci-dessous, consiste à extraire les structures contrastées de la feuille.

Extraction des structures contrastées

Dans cette étape, nous supposons que la couleur des nervures a un contraste différent de celui du limbe. Cette hypothèse est souvent vérifiée dans le cas réel. Plus la différence de contraste est élevée, plus facilement les détails du réseau de nervures peuvent être détectés. L'utilisation d'un détecteur de bord classique pourrait être envisageable pour extraire les zones de forte variance de luminosité sur l'image en niveau de gris. Cependant, nous avons remarqué qu'un tel détecteur ne capte pas certains détails du réseau de nervures. De plus, dans ce cas, ce sont essentiellement les bords des nervures qui sont repérés. En effet, quand la nervure est épaisse, son intérieur n'est pas détecté car il ne présente pas une grande variation de luminosité. Ceci est illustré dans la Figure 3.27 en fin de ce paragraphe.

Nous avons choisi l'opérateur morphologique Chapeau Haut de Forme (CHDF) (top-hat en anglais) qui remédie aux limites des détecteurs de bord. Il est appliqué sur des images d'intensité pour détecter les structures contrastées. Il produit en sortie une image en niveaux de gris, notée ici I_{CHDF} , qui contient les zones de contraste local. Une valeur élevée dans I_{CHDF} correspond à une zone de contraste élevé. Le CHDF peut être obtenu de deux manières différentes :

- En soustrayant l'image initiale de l'image de l'ouverture. Il s'agit dans ce cas du CHDF blanc (white top-hat).
- En soustrayant l'image fermée de l'image initiale. Il s'agit dans ce cas du CHDF noir (black top-hat).

Formellement, le CHDF est défini comme suit :

$$I_{CHDF} = CHDF_B(I_{TM}) = \gamma_B(I_{TM}) - I_{TM} \quad \text{ou bien} \quad I_{CHDF} = CHDF_B(I_{TM}) = I_{TM} - \varphi_B(I_{TM}) \quad (3.21)$$

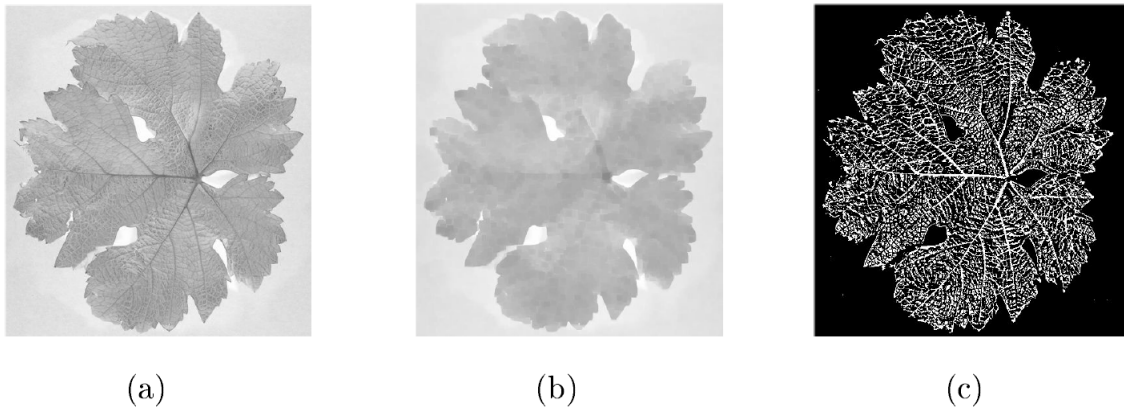


FIGURE 3.26 – Transformation du Chapeau Haut De Forme (CHDF). (a) image originale inverse d'intensité; (b) image fermée; (c) $I_{CHDF_{bin}}$: binarisation du résultat de CHDF.

Nous utilisons dans notre cas le CHDF noir car nous partons de l'hypothèse que les nervures sont plus claires que la couleur globale de la feuille. Cette supposition est vérifiée sur la grande majorité des feuilles, en particulier, sur celles de la base Flavia [WBX⁺07]. L'idée sous-jacente de la soustraction de l'image fermée de l'image initiale est d'éliminer les structures contrastées dans un premier temps avec la fermeture. Ces structures réapparaissent dans la suite en réalisant la soustraction (cf. Figure 3.26). Ainsi une zone homogène dans l'image d'intensité conduira à une zone très foncée dans l'image I_{CHDF} et *vice versa*. Le choix de l'élément structurant est encore une fois primordial. Pour faire disparaître toutes les structures contrastées, en particulier celles qui appartiennent au réseau de nervures,

il est nécessaire de choisir un élément structurant de taille suffisante pour contenir ces structures. Nous avons essayé différentes tailles : $T_B = 3$, $T_B = 5$, $T_B = 7$ et $T_B = 9$. Nous avons retenu un élément structurant carré de taille $T_B = 7$ car il donne le meilleur résultat en terme de détection des détails du réseau de nervures.

Afin de valider le choix de l'opérateur $CHDF$, comparons ses résultats avec ceux du détecteur de bord Sobel (cf. Figure 3.27). La figure montre que l'image binaire du CHDF $I_{CHDF_{bin}}$ contient tous les détails du réseau de nervures mais également un très grand nombre de faux positifs. Dans le cas de Sobel, le bruit est beaucoup plus réduit mais la détection des détails du réseau est moins précise. Les résultats sur une région de la feuille, contenant des nervures secondaires et tertiaires, illustrent clairement la différence entre les deux méthodes. Les détails de la nervation sont mieux détectés avec l'opérateur $CHDF$ et le bruit présent n'a pas une grande importance à cette étape si l'on utilise une méthode de filtrage dans la suite.

La prochaine étape consiste donc à isoler le réseau de nervures sur l'image $I_{CHDF_{bin}}$ de façon à éliminer les faux positifs et ne garder que les structures "nervures". Elle consiste en un ensemble d'opérations effectuées sur $I_{CHDF_{bin}}$.

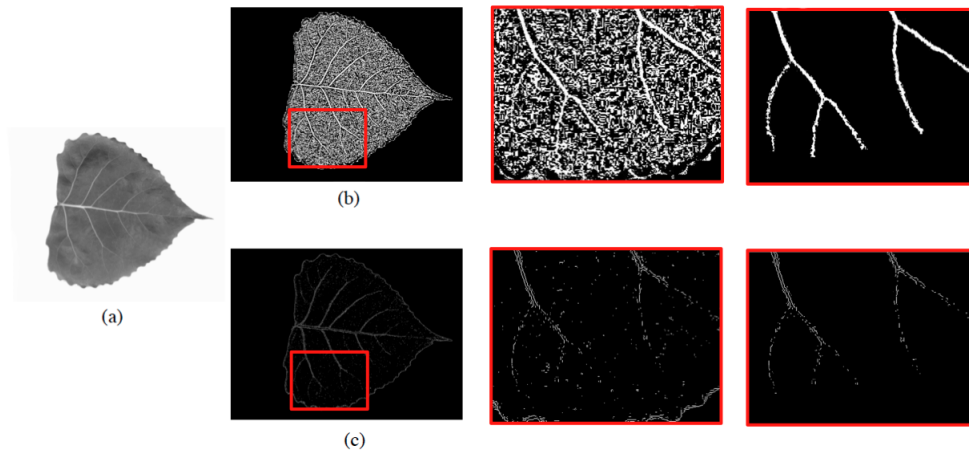


FIGURE 3.27 – Différence entre les opérateurs $CHDF$ et Sobel. (a) Image originale en niveaux de gris. (b) Structures contrastées détectées par l'opérateur morphologique Chapeau Haut de Forme (CHDF). (c) Structures contrastées détectées par le détecteur de bord Sobel.

Détection des structures linéaires

Certains types de nervations contiennent des nervures parfaitement linéaires. Citons, par exemple, la nervation dichotomique et la nervation parallèle. D'autres types de feuilles

contiennent des nervures plus courbées (comme la nervation palmée et la nervation curvinervée). Ces types sont présentés dans la figure 3.28. Dans les deux cas, nous partons de l'hypothèse que le réseau de nervures peut être représenté par une suite de structures linéaires. Le but de cette étape est la détection de ces structures qui permettent de construire l'intégralité du réseau.

L'image de départ est l'image du contraste $I_{CHDF_{bin}}$ qui, bien qu'elle soit très bruitée, comporte un maximum de détails de la nervation. Notre but est de détecter le plus possible de structures appartenant à la nervation. Le filtrage du bruit est effectué dans une étape ultérieure.

Pour détecter les structures linéaires, des ouvertures morphologiques avec un élément structurant linéaire sont utilisées. Pour pouvoir détecter des nervures dans différentes directions, il convient de faire varier l'orientation de l'élément structurant.

Dans notre algorithme, huit directions sont considérées (elles ont les angles suivants par rapport à l'horizontale : 0° , 30° , 45° , 60° , 90° , 120° , 135° , 150°). Les structures linéaires des nervures sont d'abord détectées suivant ces huit directions. Puis l'union des huit images résultantes (cf. Figure 3.29) est calculée :

$$\gamma^{union} = \bigcup_{i=1}^8 \gamma_{B_i}^i \quad (3.22)$$

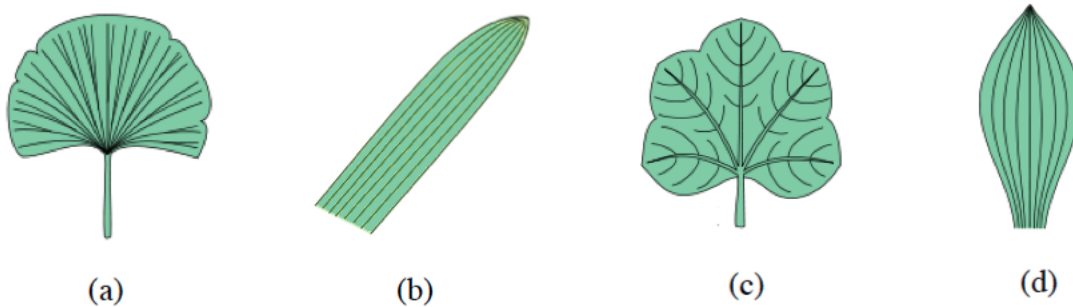


FIGURE 3.28 – Différents types de nervation (a) dichotomique (b) parallèle (c) palmée (d) curvinervée

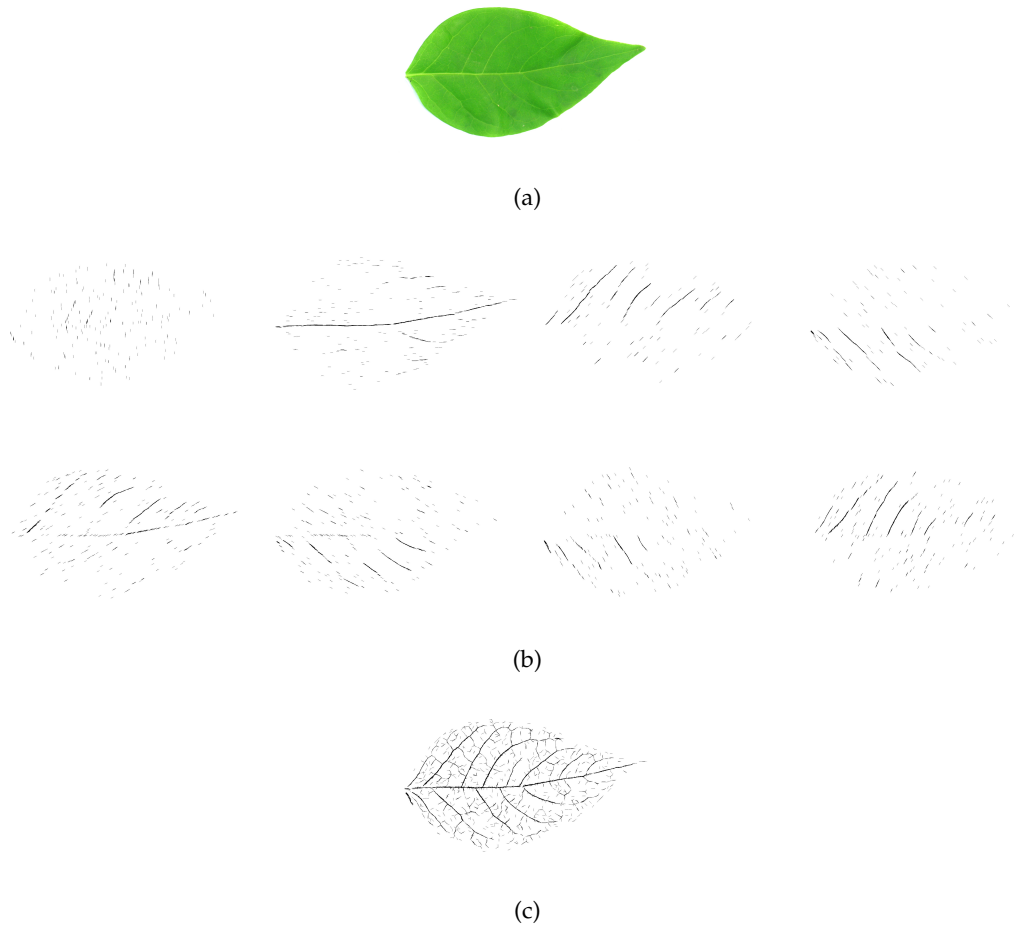


FIGURE 3.29 – Ouvertures morphologiques directionnelles avec un élément structurant linéaire de taille $T_B = 15$. (a) Image originale. (b) Ouvertures directionnelles suivant huit directions. (c) Union des huit ouvertures directionnelles.

L'image de l'union des ouvertures directionnelles est une première approximation du réseau de nervure. Elle contient des nervures mieux isolées mais également beaucoup de faux positifs. L'étape suivante est dédiée au filtrage.

Filtrage

Le but de cette étape est de garder les structures "nervures" tout en éliminant celles "non nervures". Pour cela, il est nécessaire de modéliser le bruit qui constitue l'ensemble des faux positifs. La figure 3.30 montre une détection des nervures avec des éléments structurants de taille différente. Nous remarquons un très grand nombre de faux positifs qui se réduit quand on augmente la taille de T_B . Une idée serait donc d'utiliser un élément structurant de grande taille. Cependant, l'utilisation d'un élément structurant de taille

importante est restrictive et peut empêcher l'extraction de structures appartenant aux nervures. Nous nous trouvons donc face à un problème classique de seuillage où il n'existe pas de seuil optimal global.

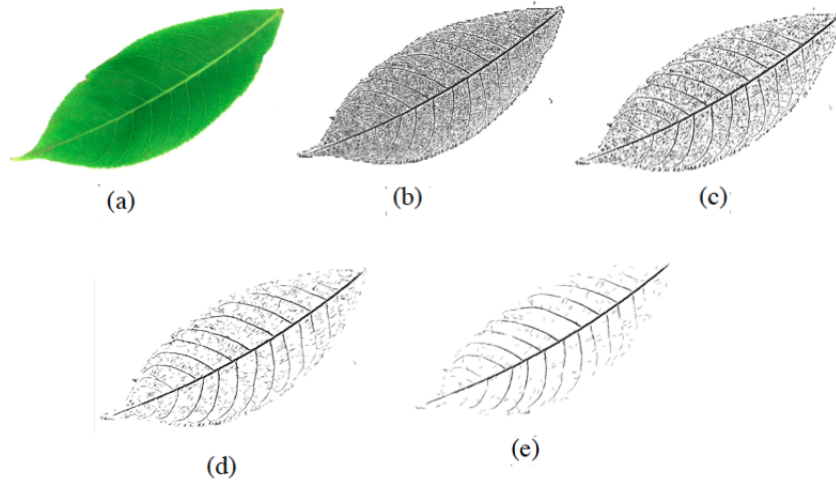


FIGURE 3.30 – Effet de la variation de la taille de l'élément structurant. (a) Image couleur originale. (b) $T_B = 3$. (c) $T_B = 7$. (d) $T_B = 11$. (e) $T_B = 15$.

Afin de résoudre le problème de seuil global, nous considérons séparément les huit images d'ouvertures directionnelles et nous cherchons à obtenir un seuil approprié pour chacune d'elles. Les images d'entrée considérées à cette étape sont les images d'ouvertures directionnelles avec un élément structurant de taille $T_B = 3$. Ce sont des images très bruitées mais qui contiennent généralement l'intégralité du réseau de nervures.

Dans une première étape, toutes les composantes connexes de chaque image sont détectées. En effet, l'ouverture détecte des structures linéaires qui, en cas d'adjacence, forment une composante connexe que nous dénommons CC. Le diamètre de chaque composante est calculé. Il correspond à la distance discrète maximale entre deux points de CC.

$$Diametre(CC) = \max\{d(p_i, p_j) | p_i, p_j \in CC\} \quad (3.23)$$

où $d(p_i, p_j)$ est la distance discrète entre les points p_i et p_j .

Le diamètre est plus significatif que la surface car l'élongation d'une composante connexe est un critère important du filtrage. En effet, plus la valeur du diamètre est élevée, plus la composante connexe CC a une chance de faire partie des nervures car le bruit est représenté généralement par des segments parasites de taille très réduite.

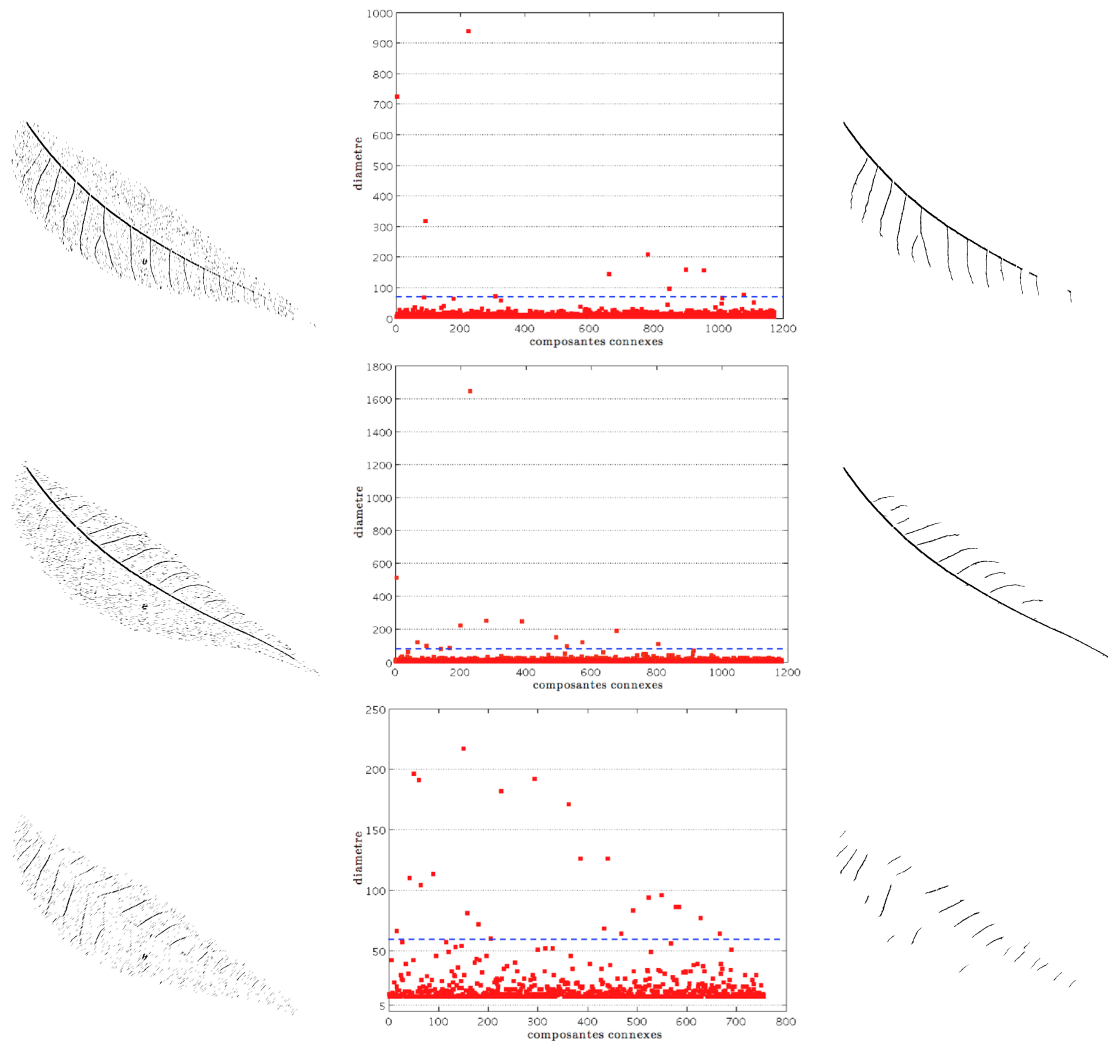


FIGURE 3.31 – Graphes des diamètres des composantes connexes pour les trois premières images d’ouvertures directionnelles. Un seuil adaptatif est calculé sur chaque graphe, il est représenté par la ligne discontinue. L’image directionnelle filtrée en se basant sur ce seuil est montrée à droite de chaque graphe.

Nous établissons une liste de diamètres des différentes composantes connexes CC pour chaque image d’ouverture directionnelle. Dans la figure 3.31, des exemples de graphes de diamètres sont montrés sur trois images. En axe des abscisses, nous avons la liste de composantes connexes correspondant à une image directionnelle donnée. Les petits carrés rouges représentent les tailles des diamètres associés à ces CC. L’analyse des graphes montre qu’il existe un très grand nombre de composantes connexes de diamètre réduit. Nous remarquons une grande densité des petits carrés rouges pour de faibles valeurs de diamètre. A l’inverse, nous observons sur les trois images des composantes connexes dont le diamètre a une valeur élevée. Elle ne peuvent en aucun cas être des faux positifs et elles appartiennent

certainement au réseau de nervures.

Donc les deux critères importants retenus sont :

- la valeur du diamètre d d'une composante connexe CC.
- la fréquence du diamètre d , en prenant en compte toutes les composantes connexes CC de l'image.

Sur la base de ces deux critères, nous définissons la méthode de filtrage comme suit.

Pour chaque image d'ouverture directionnelle I_o

Algorithme 1: Pseudo-Code de l'algorithme de filtrage des nervures

```

Données :  $I_o$  : Image de nervures non filtrées
1  $B$  : élément structurant de taille  $T_B$ 
Sorties :  $I_f$  : Image binaire de nervures filtrées
/* Initialiser un seuil  $s$  */
2  $s \leftarrow 0$ 
/* Matrice d'occurrence de diamètres */
3  $Occ \leftarrow$  nombre de composantes connexes CC de diamètre égal à  $d$ .
/* Liste de diamètres */
4  $L_d \leftarrow \{d_1, d_2, \dots, d_n\}$ 
5  $L_d \leftarrow \{L_d \text{ sans doublons}\}$ 
/* Calculer la moyenne  $moy_d$  des diamètres de la liste  $L_d$  */
6  $moy_d \leftarrow \frac{1}{card(L_d)} \sum_{i=0}^n d_i$ 
7 pour chaque  $d_i \in L_d$  faire
8   si  $Occ(d_i) > 3 \vee d_i < moy_d$  alors
9     si  $d_i > s$  alors
10     $s \leftarrow d_i$ 
/* Éliminer toutes les composantes connexes CC ayant un
   diamètre inférieur à  $s$  dans l'image  $I_o$  */
11 pour chaque  $CC_i \in I_o$  faire
12   si ( $diametre(CC_i) \leq s$ ) alors
13    $CC_i \leftarrow \emptyset$ 

```

Description de l'algorithme du filtrage

Le but de cet algorithme est d'éliminer les structures linéaires qui ont un diamètre inférieur à un seuil s . Nous avons démontré qu'il n'existe pas un seuil global ou fixe. La valeur du seuil doit donc être calculée en fonction des diamètres et de leur fréquence.

L'instruction 3 sert à initialiser la structure Occ contenant la fréquence de chaque diamètre (par exemple, $Occ(5) = 12$ s'il existe 12 composantes connexes ayant un diamètre égal à 5). Les instructions 4 à 5 visent à créer une liste de diamètres sans doublons. Si au moins une composante connexe CC a pour diamètre d , d doit figurer une seule fois dans la liste L_d .

Dans l'instruction 6, la moyenne de tous les diamètres de la liste L_d est calculée. L'étape d'élimination des doublons dans L_d est primordiale car elle garantit que la moyenne n'est pas affectée par la présence d'un grand nombre de CC de taille réduite. Les instructions de 7 à 13 présentent la fonction principale du filtrage : deux conditions sont utilisées, une sur la fréquence du diamètre et une deuxième sur sa taille. Si l'une des conditions est vérifiée, alors d pourrait être le seuil s recherché. Avec la condition sur la valeur de d , nous considérons que toutes les composantes connexes CC de taille inférieure à la moyenne moy_d sont des segments parasites. Des expérimentations ont été conduites pour étudier l'impact du remplacement de la moyenne par la valeur médiane dans la liste L_d . Les résultats étaient comparables si nous considérons l'union d'ouvertures directionnelles. L'évaluation qualitative des réseaux de nervures extraits montre une légère précision supplémentaire en calculant la moyenne moy_d . C'est pour cette raison que le calcul de la valeur moyenne est retenue lors du critère de longueur du diamètre.

La condition sur la fréquence suppose que s'il existe au moins trois CC de diamètre égal, elles sont considérées comme du bruit. Cette condition concerne un cas fréquent rencontré dans les images d'ouverture directionnelles. En effet, il existe des directions où toutes les CC sont de taille réduite et ont des diamètres très proches. Cela est dû généralement à l'absence de nervures dans cette direction (cf. Figure 3.32). La condition sur la fréquence représente alors un filtre supplémentaire dans ce cas.

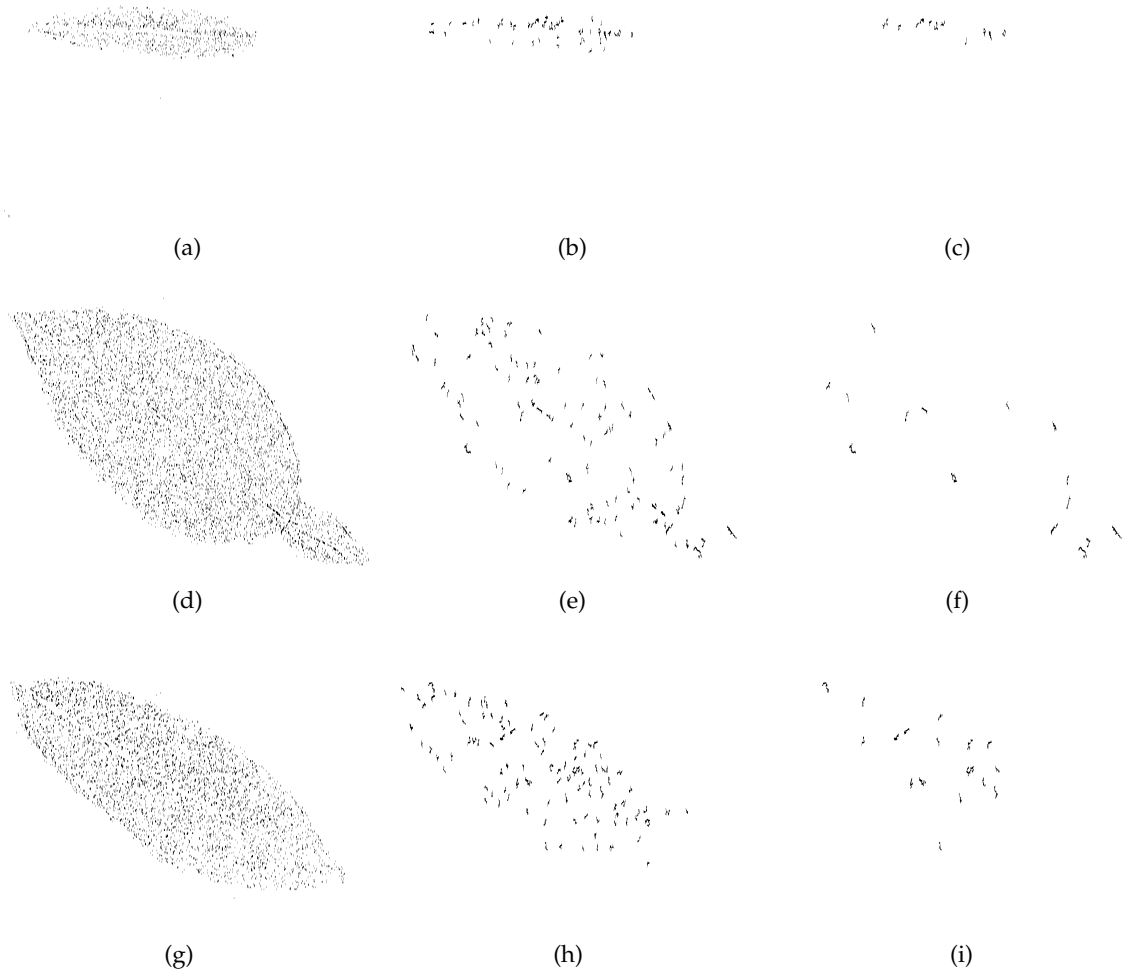


FIGURE 3.32 – Illustration des deux critères du filtrage. (a,d,g) : exemple d’images d’ouvertures directionnelles avec $T_B = 3$. (b,e,h) : filtrage avec la condition sur la longueur des diamètres. (c,f,i) : filtrage avec les deux conditions (longueur et fréquence des diamètres des CC).

La figure 3.33 montre un exemple de filtrage d’images d’ouvertures directionnelles. La différence entre l’image d’union d’ouvertures avant et après le filtrage montre que l’algorithme de filtrage élimine les faux positifs tout en gardant les structures appartenant à la nervation.

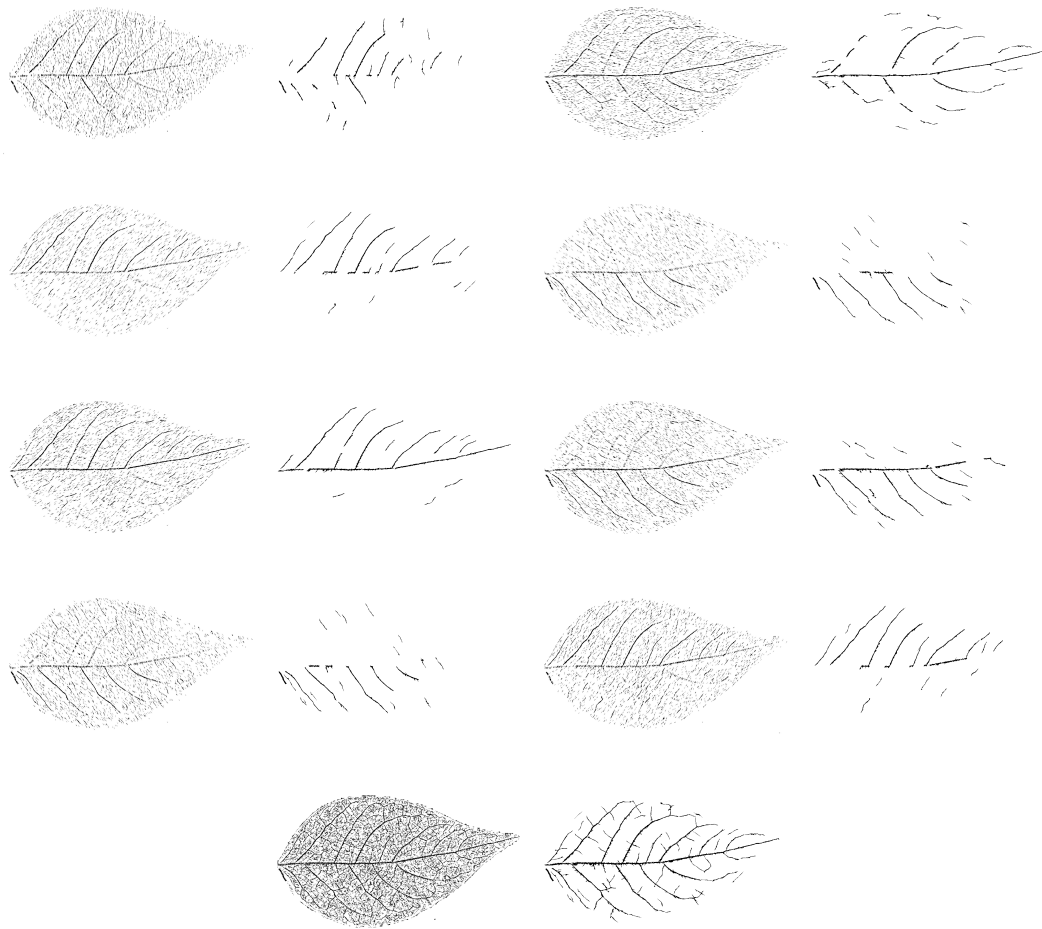


FIGURE 3.33 – Filtrage des images d’ouvertures directionnelles.

3.3.3 Résultats d’extraction des nervures

Un aperçu global des résultats d’extraction de nervures sur la base Flavia est donné dans l’annexe A. Cela montre que le réseau de nervures est généralement extrait avec une précision satisfaisante.

En revanche, la variété des feuilles de la base Flavia fait que sur certaines images, il est moins évident de repérer les nervures. Cela est dû principalement au contraste très faible, voire absent, sur certaines feuilles. Comme le montre la figure 3.34, ce phénomène peut se produire sur une partie ou sur l’intégralité du réseau de nervures et représente la cause principale des cas d’échec de notre algorithme d’extraction. En fait, la majorité des méthodes d’extraction de nervation sont testées sur des bases de feuilles privées ou sur des sous bases. En effet, il leur faut garantir une certaine qualité des images en sélectionnant celles où les nervures sont fortement contrastées.

Dans notre cas, nous avons utilisé la base publique Flavia pour l'extraction mais également pour la description du réseau de nervures.

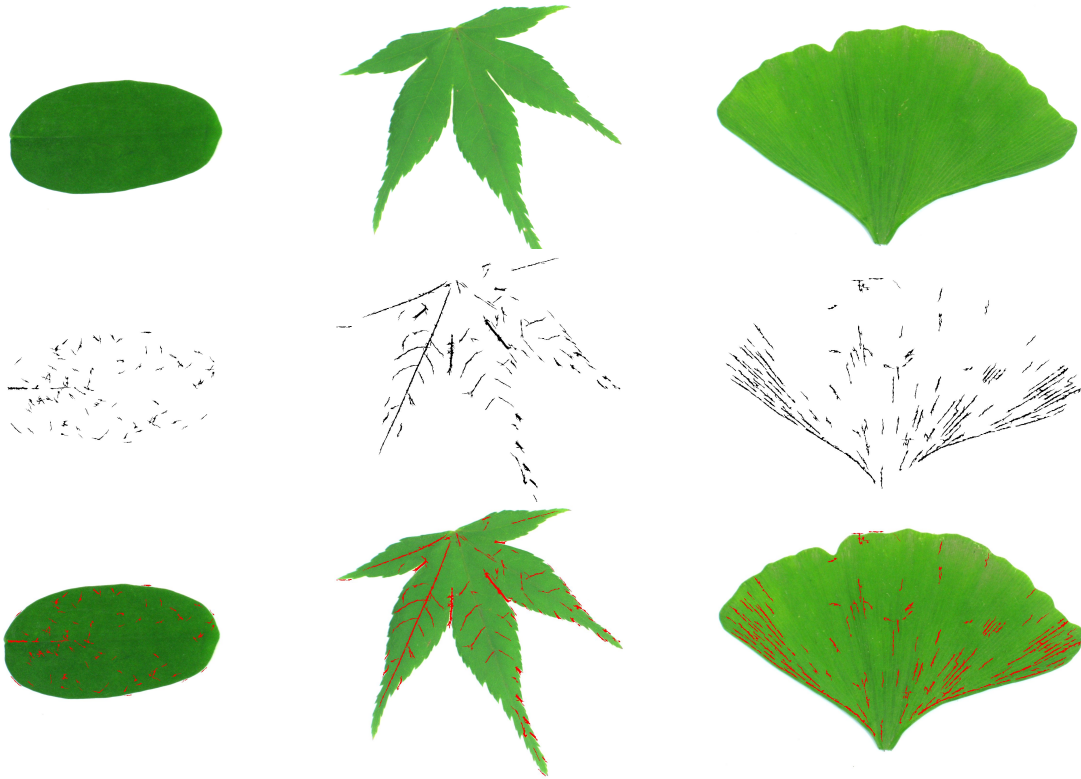


FIGURE 3.34 – Cas d'échec d'extraction de nervures. Première ligne : images originales, deuxième ligne : nervures binaires, troisième ligne : nervures reportées sur les images originales.

Dans l'absence d'une vérité terrain pour évaluer l'approche d'extraction de nervures proposée, comparons les réseaux extraits avec les résultats de la littérature.

Pour une comparaison raisonnable, nous nous basons sur les résultats de l'algorithme morphologique d'extraction de nervures proposé par Abdul Kadir et al. [KNSS11b] qui utilise des opérateurs de la morphologie mathématique. Plus précisément, les auteurs calculent la soustraction de l'image en niveaux de gris à partir d'une ouverture morphologique, ce qui revient à calculer l'opérateur $CHDF$. Ils appliquent ensuite un seuil global sur l'image I_{CHDF} afin de réduire les faux positifs. Si nous comparons cela avec notre approche, les auteurs s'arrêtent à l'étape de détection de structures contrastées. Dans notre approche, cette étape est intermédiaire et elle est suivie d'une étape de détection des structures linéaires et une étape de post-filtrage.

Examinons la figure 3.35 pour comparer nos résultats avec ceux de [KNSS11b]. Notre méthode réduit considérablement les faux positifs et obtient des nervures mieux construites.

Ceci s'explique par notre choix d'utiliser un élément structurant linéaire qui correspond mieux à la morphologie des nervures alors que, dans [KNSS11b], les auteurs ne procèdent pas à une étape de détection de structures linéaires. La deuxième différence est le fait que le processus de filtrage est presque absent dans [KNSS11b] et se résume au seuillage de l'image des structures contrastées I_{CHDF} .

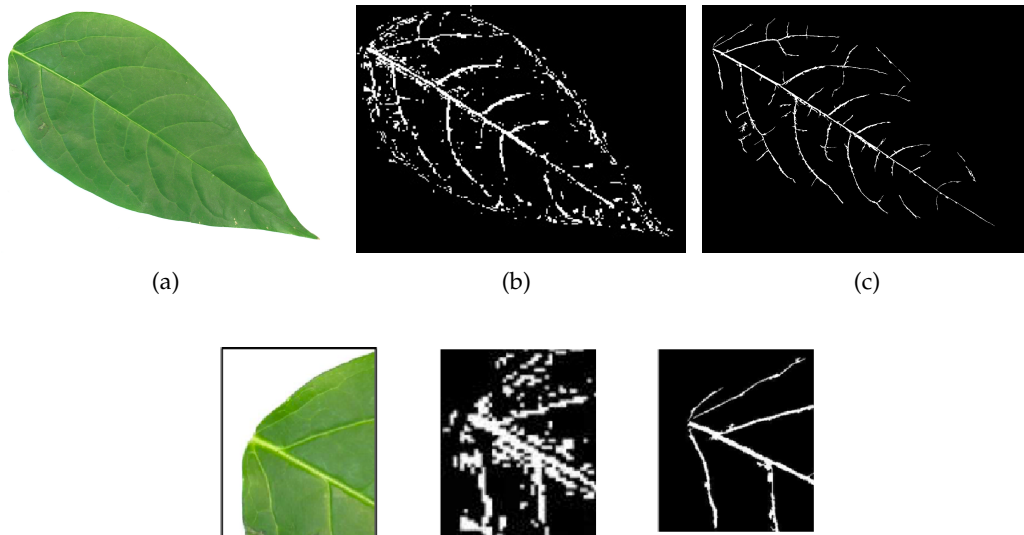


FIGURE 3.35 – Comparaison de notre approche avec la méthode morphologique d'extraction de nervures utilisée dans [KNSS11b]. (a) image couleur originale de la base Flavia ; (b) nervures extraites dans [KNSS11b] ; (c) nervures extraites par notre approche.

En fin de section, nous fournissons un pseudo-code pour l'algorithme d'extraction de nervures

3.3.4 Description de la nervation

Une fois le réseau de nervures détecté, il faut le décrire en vue d'identifier l'espèce de la feuille. L'utilisation de la nervation comme critère principal d'identification d'espèces végétales n'est pas courante dans la littérature. Nous avons présenté dans le paragraphe 2.3.3 les méthodes de description existantes. Nous rappelons que ces méthodes n'obtiennent pas, généralement, une précision élevée. De plus, les approches d'identification des feuilles via leurs nervations sont rarement évaluées sur des bases publiques (sauf [KNSS11b]). Dans ce manuscrit, nous proposons un nouveau descripteur de nervures qui comprend deux descriptions différentes des nervures :

- une détection des directions principales des nervures, non exploitées de manière assez

Algorithme 2: Pseudo-Code de l'algorithme d'extraction de nervures

Données : I : Image couleur originale
Sorties : $I_{nervures}$: Image binaire des nervures

```

1   $B$  : élément structurant de taille  $T_B$ 
   /* Calcul de l'image d'intensité du canal vert */
2   $I_{NG} \leftarrow \text{canal\_vert}(I)$ 
   /* image d'intensité inverse */
3   $I_{inv} \leftarrow \text{inverse}(I_{NG})$ 
   /* toggle mapping */
4  pour chaque  $p(x, y) \in I_{inv}$  faire
5    si  $[\gamma(I_{inv})](p) - I_{inv}(p) \leq I_{inv}(p) - [\varphi(I_{inv})](p) - I_{inv}(p)$  alors
6    |    $I_{TM}(p) = [\gamma(I_{inv})](p)$ 
7    sinon
8    |    $I_{TM}(p) = [\varphi(I_{inv})](p)$ 
   /* Chapeau haut de forme */
9   $I_{CHDF} \leftarrow (I_{TM}) - \gamma_B(I_{TM})$ 
   /* Binariser l'image du chapeau haut de forme */
10  $I_{CHDF_{bin}} \leftarrow \text{Binariser}(I_{CHDF})$ 
   /* Ouvertures directionnelles et filtrage */
11 pour  $i \leftarrow 1$  a 8 faire
12 |    $I^i \leftarrow \gamma_{B_i}^i(I_{CHDF_{bin}})$ 
13 |    $I_{filtree}^i \leftarrow \text{filtrage}(I^i)$ 
   /* image binaire de nervures */
14  $I_{nervure} \leftarrow \bigcup_{i=1}^8 I_{filtree}^i$ 

```

profonde dans la littérature.

- une description spatiale des points de nervation.

Dans les deux sections suivantes, ces deux descriptions sont présentées.

Description des directions principales des nervures

Dans cette partie, nous cherchons à représenter le réseau de nervures par un ensemble de structures linéaires (segments ou droites). Pour ce faire, notre choix s'est porté sur la transformée de Hough qui s'avère utile pour la détection de droites dans une image et qui présente des avantages dans notre cas particulier. En effet, il existe sur les réseaux de nervures extraits un certain nombre de faux positifs et de nervures discontinues. La transformée de Hough se caractérise par une bonne insensibilité au bruit et permet de détecter des objets présentant des petites discontinuités ou étant partiellement occultés.

Rappelons tout d'abord le fondement théorique de la transformée de Hough. Soit \mathcal{D} une droite dans un repère (X, Y) qui a pour équation $y = ax + b$ qui s'écrit d'une façon équivalente $-b = ax - y$. Le principe consiste à projeter les coordonnées (x, y) dans un espace de paramètres (A, B) appelé aussi domaine de Hough. La droite \mathcal{D} est représentée par un point (a, b) dans l'espace des paramètres. Notons également que les points $(x_1, y_1), \dots, (x_n, y_n)$ d'une même droite $\mathcal{D} : y = ax + b$ seront représentés par n droites se coupant en un point (a, b) dans l'espace des paramètres.

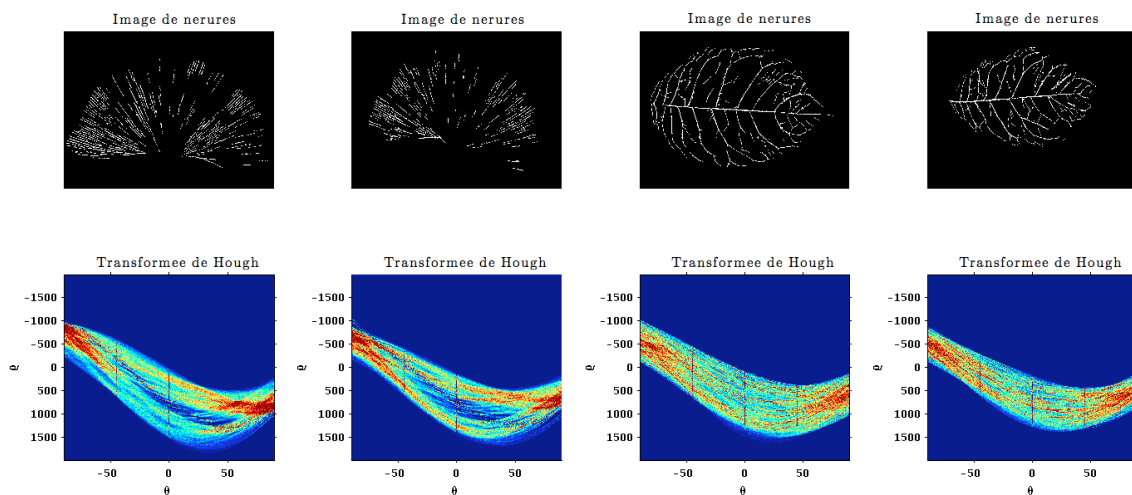


FIGURE 3.36 – Transformée de Hough sur des images de nervures.

La figure 3.36 montre le résultat de l'application de la transformée de Hough sur des images de nervures. Les deux premières images sont calculées à partir de feuilles de la même espèce tandis que les deux dernières images correspondent à une autre espèce végétale. Notons la similarité de la description de Hough pour les images de même classe et la différence des signatures des images de classe différente. Cela confirme que la méthode de Hough, appliquée aux nervures, est capable de fournir des signatures similaires pour les images de même espèce.

Cependant, la transformée de Hough est, par définition, sensible à la rotation (cf. Figure 3.37). Cette étape nécessite un alignement qui peut se faire soit via un ACP soit en repérant l'apex et la base pour identifier la direction principale de la feuille. En examinant les images de la base de test Flavia que nous allons utiliser pour évaluer cette approche, nous avons remarqué que, dans chaque classe, il existe suffisamment d'images qui ont le même alignement. Donc, la sensibilité à la rotation de la description de Hough aura peu d'impact surtout si nous considérons une faible valeur de k , en examinant les k plus proches voisins de chaque image requête. Par contre, cela aura un impact sur le calcul de la MAP (Mean Average Precision) puisque nous considérons toutes les images retournées ($k = Nb$ avec Nb le nombre d'images dans la base).

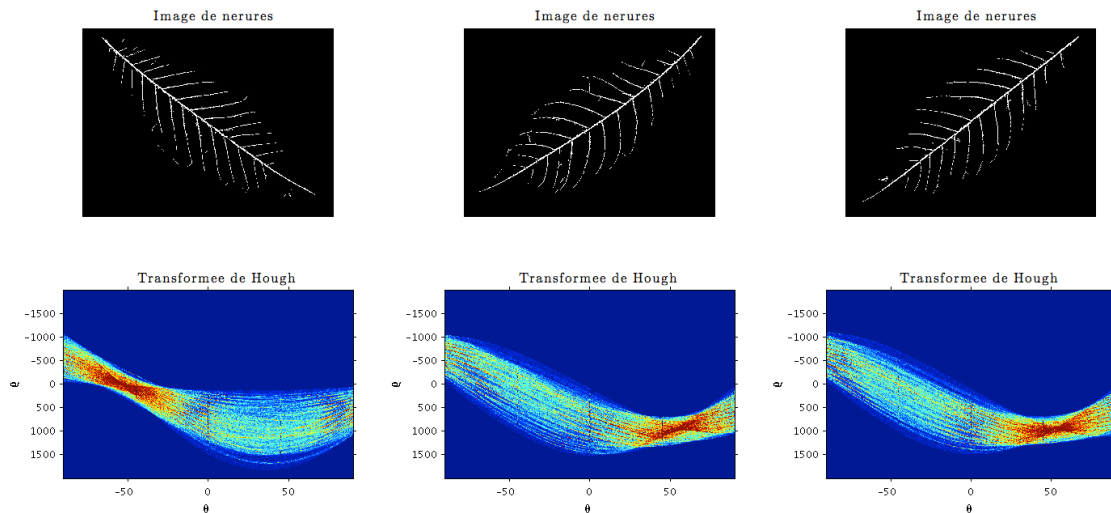


FIGURE 3.37 – Sensibilité à la rotation de la transformée de Hough

Description spatiale des nervures

Dans la littérature, la description spatiale des nervures est généralement exprimée à l'aide des distances entre les points de jonction et les points de terminaison [PHN06,

PHN08, NHK08]. Nous rappelons que ce type de description est fortement dépendant de la qualité des nervures extraites et n'est pas adéquat pour décrire un réseau déconnecté ou bruité.

Dans notre approche, nous cherchons à décrire les relations spatiales entre les points de nervures en utilisant un des scénarios du contexte de forme étendu présenté dans le paragraphe 3.2.3. Deux schémas sont proposés en faisant varier les points de l'ensemble votant :

ASC_{PN-PN} **Relations spatiales entre les points nervures.**

Ce scénario vise à décrire l'arrangement spatial des nervures via un contexte de forme (cf. Figure 3.38) . L'ensemble de calcul et l'ensemble votant sont identiques et contiennent les points de nervation. Cette description peut être utile pour une classification de feuilles selon le type de nervation.

$$\mathcal{C} = \mathcal{V} = PN = \{\text{points de nervation}\}$$

ASC_{PN-PC} **Corrélation spatiale entre le réseau de nervures et le contour.**

Dans le scénario précédent, deux réseaux de nervures ayant un arrangement spatial proche auront des descripteurs similaires. Nous souhaitons apporter plus de capacité de discrimination en intégrant l'information contour. Nous suggérons à travers ce scénario que l'ensemble votant soit composé des points du contour, c'est à dire, il s'agit de décrire la position des nervures dans un contexte défini par le contour (cf. Figure 3.38). Les ensembles de calcul et de vote sont alors distincts et se définissent comme suit :

$$\mathcal{C} \neq \mathcal{V}, \mathcal{C} = PN = \{\text{Points de nervation}\} \text{ et } \mathcal{V} = PC = \{\text{Tous les points du contour}\}$$

Ce scénario est un cas particulier du scénario *ASC_{PS-PC}* défini dans 3.2.3. Dans ce cas, les points saillants de la feuille sont représentés exclusivement par les points de nervures.

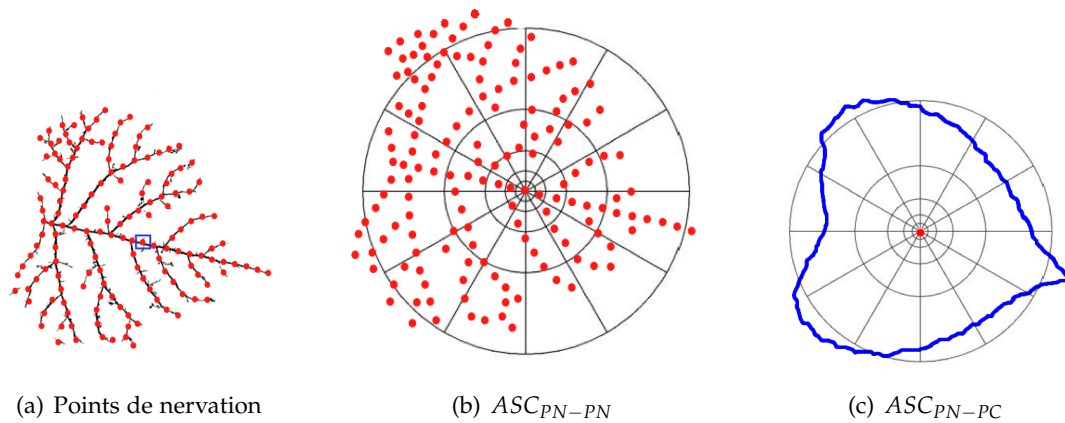


FIGURE 3.38 – Scénarios de description de nervures.

3.4 CONCLUSION

Dans ce chapitre, nous avons introduit différentes méthodes de description d'une feuille pour permettre l'identification de son espèce végétale. Nous avons considéré trois caractères foliaires : la forme, les points saillants et la nervation. Les méthodes d'extraction et de description de chaque caractère sont résumées dans le tableau 3.3

Caractère	Méthode d'extraction	Méthode de description
Forme	Seuillage d'Otsu	Description triangulaire multi-échelle (<i>TSL</i> , <i>TSLA</i> et <i>TOA</i>)
Points saillants	Détecteur de Harris	Contexte de forme étendu <i>ASC</i>
Nervation	Méthode morphologique de détection des structures linéaires	Description des directions principales du réseau de nervures (<i>Hough</i>) et description des relations spatiales entre les points nervures

TABLE 3.3 – Résumé des approches proposées

Pour toutes les descriptions proposées de forme et de points saillants, nous avons utilisé une approche de mise en correspondance locale décrite dans 2.8.3, page 46.

Le chapitre suivant sera consacré à l'évaluation de ces différentes méthodes de description et à l'étude de l'impact de chaque caractère dans l'identification des espèces végétales.

Chapitre 4

Evaluation

Nous présentons dans ce chapitre les évaluations des descripteurs proposés sur différentes bases d'images de feuilles accessibles au grand public.

Swedish [Söd01] (cf. Paragraphe 4.1.1) est la première base de feuilles publiée et partagée. Elle est utilisée dans la littérature pour l'évaluation des descripteurs génériques de forme et contient des images de feuilles simples et composées.

La base Flavia [WBX⁺07] (cf. Paragraphe 4.2.1) contient des images de bonne qualité contenant des feuilles simples sans pétioles. Elle est surtout utilisée pour l'évaluation des approches d'identification d'espèces végétales.

La base One hundred species (Leaf100) [MCO13] (cf. Paragraphe 4.3.1) comporte un nombre plus important d'espèces. Elle contient uniquement des masques d'images binaires. Par conséquent, nous n'avons évalué que les descripteurs de forme sur cette base.

Nous avons également utilisé les bases fournies par la campagne d'évaluation ImageCLEF en 2011, 2012 et 2013. Ces bases contiennent des images de feuilles très variées appartenant à trois catégories différentes : scan, pseudo-scan et photographie (cf. paragraphe 4.4.1). Elles sont enrichies chaque année par de nouvelles images et de nouvelles espèces. Les bases ImageCLEF sont plus proches du cas d'utilisation réel, en plus, elles représentent les bases d'apprentissage du système d'identification utilisé dans le projet Pl@ntNet.

4.1 ÉVALUATION SUR LA BASE SWEDISH

4.1.1 Caractéristiques de la base Swedish

La base Swedish [Söd01] contient 1125 images uniformément réparties en 15 classes. Chaque espèce est représentée par 75 images (cf. Figure 4.1). Dans la base Swedish, deux classes contiennent des feuilles composées tandis que les autres classes comportent des feuilles simples.

Notons la grande similarité entre les feuilles des classes #1, #3 et #9. La principale difficulté de cette base est de différencier ces classes.

La base Swedish a été largement utilisée : dans les approches génériques de description de forme [BCF⁺08, WFB⁺14] et dans les méthodes d'identification d'espèces [RWZ12, ZHW12, WG13].

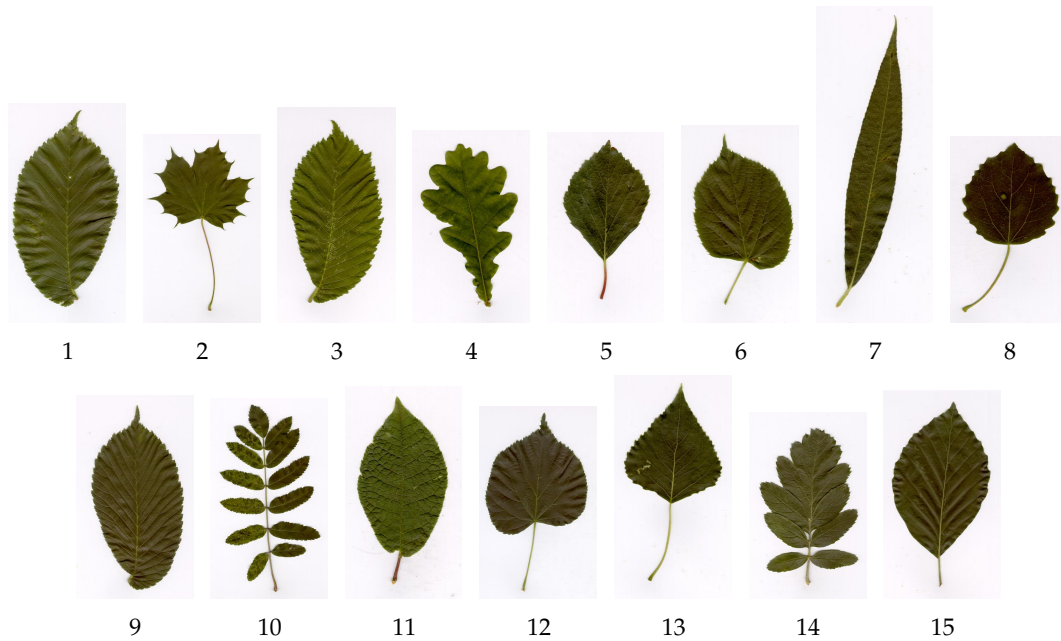


FIGURE 4.1 – La base Swedish. Une image par espèce est montrée.

Protocole d'évaluation

Le protocole d'évaluation commun utilisé sur Swedish consiste à diviser aléatoirement chaque classe en un ensemble d'apprentissage composé de 25 images et un ensemble de test contenant les 50 images restantes. Ici, la précision à $Knn = 1$ est calculée : nous vérifions si la première image retournée appartient à la même espèce que l'image requête.

4.1.2 Résultats des descripteurs de forme sur la base Swedish

Étude de paramètres

Rappelons tout d'abord les paramètres des descripteurs triangulaires proposés :

- N : le contour est initialement quantifié en N points
- n : définit le voisinage d'un point de contour p_i en terme de nombre de points de contour sélectionnés de chaque côté de p_i
- $d_m^{lin/log}$: fonction de distance où m désigne la distance entre triangles successifs
- N_s : nombre de triangles (d'échelles) utilisé pour décrire chaque point du contour

Le premier paramètre étudié est le nombre de points échantillonnés sur le contour N . La précision de la première image retournée est calculée pour chaque valeur de N . Toutes

les échelles sont considérées dans cette expérimentation ($n = \frac{N}{2}-1, d_1^{lin}$). Les résultats sont montrés dans le tableau 4.1.

Nous remarquons que la précision de *TSL*, *TSLA* et *TOA* augmente avec le nombre de points du contours utilisés. Cela est surtout visible pour le descripteur *TSLA* dont la précision passe de 87.73 avec 200 points à 95.20 à 400 points. Cette hypothèse n'est pas vérifiée pour la description *TAR* qui obtient la meilleure précision à 200 points mais la précision à 400 points est assez proche.

N	50	100	200	400
<i>TSL</i>	80.13	83.73	87.73	95.20
<i>TSLA</i>	88.66	91.43	92.88	94.08
<i>TOA</i>	84.66	86.13	90.80	92.00
<i>TAR</i>	85.33	86.53	89.86	88.93

TABLE 4.1 – Variation de N

Nous retenons dans la suite 400 points échantillonnés sur le contour $N = 400$. Intéressons-nous maintenant à étudier le paramètre n qui définit la longueur de la portion du contour utilisée pour décrire un point du contour p_i . Dans un premier temps, nous gardons une distance linéaire entre les triangles successifs avec différentes valeurs de n . La portion du contour correspondante à n est décrite respectivement avec cinq et dix triangles (cf. Tableau 4.3).

$N_s = 10$					
Descripteur	$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 199$
<i>TSL</i>	43.86	77.63	92.13	93.46	91.86
<i>TSLA</i>	40.93	78.66	93.33	94.13	91.73
<i>TOA</i>	44.26	80.19	92.80	94.26	91.20
<i>TAR</i>	49.46	64.00	80.40	87.17	
$N_s = 5$					
Descripteur	$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 199$
<i>TSL</i>	28.53	77.06	92.26	92.13	90.00
<i>TSLA</i>	29.06	80.66	94.53	93.73	91.20
<i>TOA</i>	39.20	85.46	92.13	92.93	89.33
<i>TAR</i>	42.80	61.20	80.93	82.53	79.86

TABLE 4.2 – MAP sur Swedish suivant n et N_s

d	d_2^{lin}				d_2^{log}			
n	16	32	64	128	16	32	64	128
N_s	8	16	32	64	4	5	6	7
<i>TSL</i>	69.46	87.20	94.80	95.20	61.46	86.80	95.60	95.73
<i>TSLA</i>	71.46	88.40	94.26	94.93	64.66	89.46	96.53	96.33
<i>TOA</i>	76.26	87.33	94.80	95.20	75.46	92.40	96.00	94.26

TABLE 4.3 – MAP sur Swedish avec d_2^{lin} et d_2^{log}

D'après le tableau 4.3, différents paramètres sont retenus pour chaque descripteur :

- *TSL* : $n = 128$, d_2^{log} ce qui correspond à 7 échelles ($N_s = 7$).
- *TSLA* : $n = 64$, d_2^{log} ce qui correspond à 6 échelles ($N_s = 6$).
- *TOA* : $n = 64$, d_2^{log} ce qui correspond à 6 échelles ($N_s = 6$). Le score obtenu par *TOA* avec ces paramètres est légèrement meilleur au résultat de *TOA* publié dans [MYVB13].

Comparaison des descripteurs de forme sur Swedish

Reportons les meilleurs scores de *TSL*, *TSLA*, *TOA* et *TAR* dans le tableau 4.4. Le descripteur *TSLA* obtient le meilleur score (96.53) comparé aux autres descripteurs triangulaires. En deuxième position, nous trouvons *TSL* et *TOA* devant le descripteur *TAR*.

Méthode	Précision à $knn = 1$ (%)
<i>TSL</i>	95.73
<i>TSLA</i>	96.53
<i>TOA</i>	96.00
<i>TAR</i>	90.40

TABLE 4.4 – Performance des descripteurs triangulaires de forme sur la base Swedish

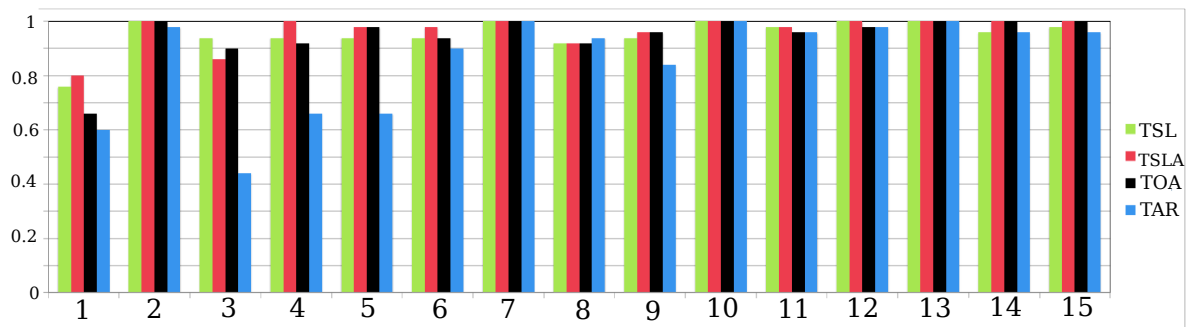


FIGURE 4.2 – Résultats des descripteurs triangulaires par classe sur la base Swedish.

La précision par classe est présentée dans la figure 4.2. Notons que 100% d'identification

correcte est enregistrée par un ou plusieurs descripteurs sur certaines classes (classes #2, #7, #10, #12, #13, #14 et #15). Nous pouvons remarquer également une légère supériorité du descripteur *TSLA* face aux autres descripteurs sur un certain nombre de classes.

Comme nous l’avons remarqué dans le paragraphe 4.1.1, la difficulté de la base Swedish réside dans les problèmes de similarité inter-classe des classes #1, #3 et #9. Cela est confirmé par la figure 4.2 qui montre que les erreurs d’identification se produisent principalement au niveau de ces classes. Pour la classe #1, considérée comme la classe la plus difficile, *TSLA* obtient le meilleur score, suivi de *TSL* et *TOA*.

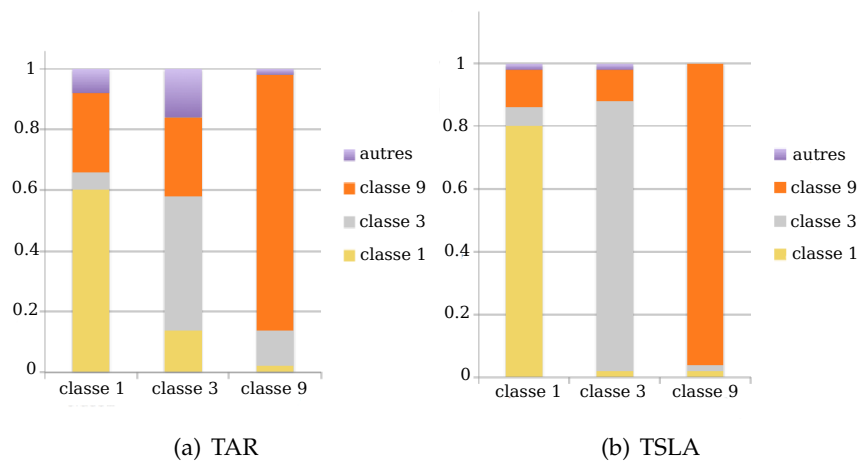


FIGURE 4.3 – Graphe de confusion entre les classes #1, #3 et #9

Nous avons calculé des graphes de confusion pour les trois classes #1, #3 et #9 afin d’analyser les cas d’identification erronée (cf. Figure 4.3). Nous comparons à travers ces graphes la robustesse des descriptions *TAR* et *TSLA* face à la similarité de forme inter-classe.

Si nous examinons les deux graphes, il est évident qu’une grande partie des mauvaises classifications est due à la similarité élevée entre les classes #1, #3 et #9. Dans le cas du descripteur *TAR*, une partie non négligeable d’erreurs de classification provient également de la confusion avec d’autres classes. Il est clair sur les deux graphes que *TSLA* traite le cas de similarité inter-classes d’une façon beaucoup plus robuste que *TAR*. Par exemple, dans le cas de la classe #3, *TSLA* réduit considérablement le taux de confusion avec les classes #1 et #9. Cela est vérifié sur les trois classes.

Identification des feuilles composées

Étudions les résultats des descripteurs triangulaires sur les feuilles composées de la base Swedish, à savoir sur les deux classes #10 et #14. Les deux descripteurs *TSLA* et *TOA* obtiennent 100% d'identification correcte sur ces deux classes en se basant sur la précision de la première image retournée. Les descripteurs *TSL* et *TAR* obtiennent respectivement des précisions moyennes (entre les deux classes) de 96.28% et 96.13%.

Dans ces deux classes, les formes globales sont différentes (cf. Figure 4.4), ce qui est confirmé par les sous résultats obtenus par nos descripteurs.

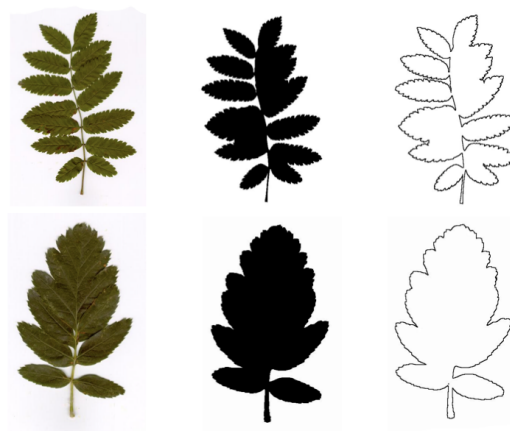


FIGURE 4.4 – Segmentation des images de feuilles composées de la base Swedish. Première ligne : image de feuille de la classe #10. Deuxième ligne : image de feuille de la classe #14.

Comparaison entre la description originale et la description locale de *TAR*

Dans la définition originale de la méthode *TAR*, Alajlan et al. [ARKF07] utilisent toutes les échelles possibles pour décrire un point du contour. La méthode DSW (Dynamic Space Wrapping) est utilisée pour la comparaison des formes. Cette approche a été évaluée sur la base Swedish par Wang et Gao [WG13]. Les auteurs utilisent un score différent de celui communément utilisé sur cette base. Ils calculent le score Bullseye défini comme suit :

Chaque image de la base est utilisée comme image requête et pour chaque réponse, le nombre d'images de même espèce que l'image requête est calculé sur les $2 \times M$ images retournées. M étant le nombre d'images dans chaque classe. Le score Bullseye est le rapport du nombre total d'images pertinentes par le nombre maximal possible qui est égal à $M \times Nb$ où Nb désigne le nombre d'images dans la base. Formellement, le score Bullseye $S_{Bullseye}$

est défini de la manière suivante :

$$S_{Bullseye} = \frac{\sum_{i=1}^{Nb} C_i}{M \times Nb} \quad (4.1)$$

Nb : nombre d'images dans la base (1125 pour Swedish). M : nombre d'images par classe (75 pour Swedish). C_i : pour la requête numéro i , R_i désigne le nombre d'images de même espèce que l'image requête, sur les $2 \times M$ images retournées. Nous suivons ce même protocole d'évaluation pour pouvoir se comparer avec le résultat dans [WG13]. Le même nombre de points échantillonnés sur le contour est également utilisé ($N = 256$). D'une façon similaire à Alajlan et al. [ARKF07], toutes les échelles sont utilisées ce qui correspond à $n = N_s = 127$ et une distance d_1^{lin} entre deux triangles successifs. Le temps moyen par requête est également mesuré. Les résultats sont reportés dans le tableau 4.5.

Méthode	TAR + DSW [WG13]	TAR+méthode de mise en correspondance locale
Score Bullseye	84.23	79.06
Temps moyen par requête (ms)	28234	175

TABLE 4.5 – Comparaison entre la mise en correspondance globale [ARKF07] et locale de la description TAR

Méthode	TSL	TSLA	TOA	TAR
Score Bullseye	85.92	87.73	86.52	83.97

TABLE 4.6 – Comparaison entre la mise en correspondance globale [ARKF07] et locale de la description TAR

La description originale TAR obtient un meilleur score (84.23 contre 79.06). Cela peut s'expliquer par le fait que la méthode DSW compare dans un premier temps la signature TAR d'un point de contour à tous les autres signatures à une même échelle. Ensuite, la distance minimale est calculée dans la table DSW pour retrouver les formes similaires. Cette méthode peut s'avérer plus précise que la mise en correspondance locale que nous utilisons où deux signatures TAR sont comparées tenant compte de toutes les échelles. De plus, Alajlan et al. [ARKF07] augmentent la capacité de distinction entre les formes de classes différentes par la définition d'une distance de dissimilarité qui utilise des descripteurs morphologiques globaux. L'ajout de ces descripteurs a amélioré le score de classification sur les bases de forme génériques telles que MPEG-7 [LLE00] et Kimia [SCTK98, SKK04].

Cependant, la méthode DSW est très coûteuse en temps de calcul. Si nous comparons le temps moyen par requête entre la mise en correspondance locale et globale, la différence

est considérable. Dans notre cas, le temps moyen par requête est réduit 161 fois par rapport l'approche TAR originale. En terme d'efficacité (précision + temps de calcul), notre approche est globalement plus efficace.

Notons que les descripteurs TSL , $TSLA$ et TOA obtiennent des scores meilleurs que celui du TAR original avec un temps de calcul considérablement plus réduit.

4.1.3 Résultats du descripteur de contexte de forme étendu

Considérons les scénarios définis dans le contexte de forme étendu : ASC_{PC-PC} , ASC_{PS-PS} , ASC_{PS-PC} et $eASC_{PS-PC}$ sur la base Swedish. La performance de ASC_{PC-PC} est évaluée en faisant varier le nombre N de points de contour et les autres scénarios sont évalués en faisant varier le nombre N_{ps} de points saillants.

N	50	100	200	400
ASC_{PC-PC}	85.60	88.80	91.46	91.73
N_{ps}	50	100	200	400
ASC_{PS-PS}	63.45	78.31	80.05	57.29
ASC_{PS-PC}	87.95	92.36	93.57	91.03
$eASC_{PS-PC}$	81.65	88.75	91.03	90.36

TABLE 4.7 – Précision ($knn = 1$) du contexte de forme étendu sur la base Swedish . N est le nombre de points échantillonnés du contour. N_{ps} est le nombre de points saillants.

En analysant les résultats du tableau 4.7, nous tirons les remarques suivantes :

- La précision du scénario ASC_{PC-PC} s'améliore quand le nombre de points du contour augmente. Cependant, le gain du passage de 200 pts à 400 pts dans la description de contour est faible (91.73 contre 91.46).
- Le fait de séparer l'ensemble de calcul et l'ensemble votant dans le scénario ASC_{PS-PC} améliore les résultats du scénario ASC_{PC-PC} où l'ensemble de calcul et l'ensemble votant sont identiques et sont composés des points du contour. L'amélioration est nette surtout quand 100 points saillants sont utilisés (92.36 contre 88.80).
- Pour tous les scénarios utilisant les points saillants, le nombre optimal de points de Harris est 200. La meilleure précision des trois scénarios est obtenue lorsque $N_{ps} = 200$. Cela confirme notre observation initiale qui constate qu'à partir d'un certain rang les points de Harris détectés représentent un bruit, c'est à dire leur localisation est moins pertinente dans le contexte d'identification d'espèces. La figure 4.5 montre l'évolution du nombre de points saillants détectés N_{ps} sur une image de feuille de

la base Swedish. Il est clair que lorsque 400 points saillants sont utilisés, un grand nombre de points sont localisés dans des endroits pas toujours pertinents.

- Le scénario $eASC_{PS-PC}$ n'améliore pas le scénario ASC_{PS-PC} en terme de précision. Cependant, le scénario $eASC_{PS-PC}$ améliore la valeur de MAP obtenue par ASC_{PS-PC} (0.773 contre 0.692).
- Le scénario ASC_{PS-PS} n'atteint pas une précision élevée par rapport aux autres scénarios. Cela montre l'importance et la robustesse du contour dans la description de la feuille.

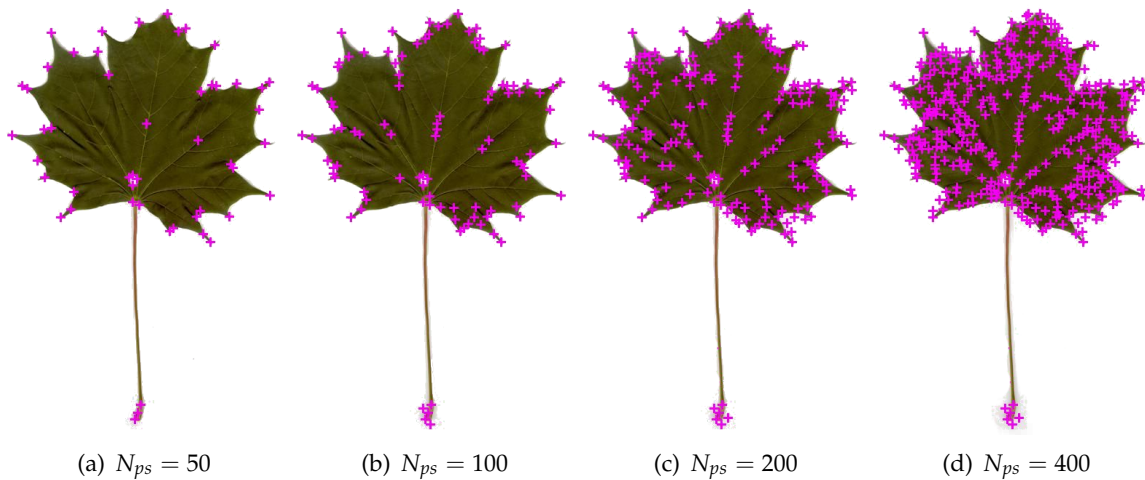


FIGURE 4.5 – Évolution du nombre de points saillants détectés N_{ps} sur une image de feuille de la base Swedish.

Comparaison entre la mise en correspondance locale et globale de la description Contexte de forme

Nous nous intéressons dans ce paragraphe à l'analyse des résultats des descripteurs contexte de forme dans les deux cas où :

- une méthode de mise en correspondance globale est utilisée
- une méthode de mise en correspondance locale est utilisée

Pour ce faire, basons-nous sur les résultats fournis dans la littérature sur la base Swedish. Le descripteur contexte de forme a été évalué dans [LJ07]. La méthode de mise en correspondance utilisée est globale et basée sur la programmation dynamique. Dans ASC_{PC-PC} , nous présentons une méthode de description identique mais elle est associée à une méthode de mise en correspondance locale. Pour la comparaison, nous avons utilisé des ensembles

d'apprentissage et de test identiques. Les résultats obtenus par les deux approches sont présentés dans le tableau 4.8.

Méthode	Précision à $knn = 1$ (%)
SC + DP [LJ07]	88.12
ASC_{PC-PC}^{128}	89.86
ASC_{PC-PC}^{200}	91.46

TABLE 4.8 – Comparaison entre la mise en correspondance locale et globale via les méthodes SC+DP et ASC_{PC-PC}

Lorsque le même nombre de points échantillonnés sur le contour ($N = 128$) est utilisé, ASC_{PC-PC} dépasse légèrement SC + DP. Si $N = 200$ dans ASC_{PC-PC} , l'écart en terme de précision est plus élevé (91.46 contre 88.12). Cela montre d'une part, que le "matching" local obtient des résultats comparables ou supérieurs à ceux du descripteur contexte de forme lorsqu'il est associé à un "matching" global. De plus, le véritable gain ne se situe pas seulement au niveau de la précision mais surtout au niveau de la performance en temps de calcul. La programmation dynamique est coûteuse en temps CPU alors que la méthode de comparaison locale utilisée dans notre cas est largement plus rapide. Cela est extrêmement important dans le cas où les méthodes d'identification sont implémentées sur des appareils embarqués.

4.1.4 Comparaison avec l'état de l'art

Nous retenons dans ce paragraphe les meilleurs résultats obtenus lors de l'étude de paramètres réalisée sur Swedish.

Méthode	Entrée	Score d'identification
<i>TSL</i>	contour	95.73
<i>TSLA</i>	contour	96.53
<i>TOA</i>	contour	96.00
ASC_{PC-PC}^{400}	contour	91.73
ASC_{PS-PS}^{200}	image couleur	80.05
$ASC_{PS_{200}-PC}$	image couleur	93.57
$eASC_{PS_{200}-PC}$	image couleur	91.03
sPACT [WR08]	image en niveaux de gris	97.92
Shape Tree [FS07]	contour	96.28
SPTC+DP ¹	image en niveaux de gris	95.33
IDSC + DP ¹	contour	94.13
IDSC + learned distance [BYL ⁺ 10]	contour	93.80
sPACT (sur contour) [WR08]	contour	90.77
Fourier ¹	contour	89.60
SC + DP ¹	contour	88.12
Söderkvist [Söd01]	contour	82.40

TABLE 4.9 – Comparaison des résultats avec l'état de l'art sur la base Swedish

Examinons le tableau 4.9. Nous distinguons les approches de description de forme dont l'entrée est le contour et les autres approches dont l'entrée est l'image intégrale (en niveaux de gris ou couleur). Cette deuxième catégorie contient généralement des méthodes qui utilisent l'information texture.

En analysant les résultats du tableau 4.9, nous remarquons que le descripteur *TSLA* est le seul qui dépasse toutes les méthodes d'identification basée sur la forme mais il est positionné deuxième considérant toutes les méthodes. Le meilleur score est obtenu par l'approche sPACT [WR08] qui représente les structures locales de la feuille. Les descripteurs *TOA* et *TSL* obtiennent le troisième et le quatrième meilleurs scores derrière *TSLA* et la méthode Shape Tree [FS07] dans la catégorie des descripteurs de forme.

Le meilleur scénario dans le contexte de forme étendu ASC_{PS-PC} obtient une performance comparable à la méthode IDSC + learned distance [BYL⁺10]. Ce scénario dépasse le descripteur SC + DP et Fourier¹.

Les résultats obtenus sur la base Swedish montrent la pertinence des descripteurs proposés particulièrement les descripteurs triangulaires de forme. Ces résultats sont à confirmer sur d'autres bases de feuilles contenant plus d'images et d'espèces.

4.2 ÉVALUATION SUR LA BASE FLAVIA

4.2.1 Caractéristiques de la base Flavia

La base d'images Flavia [WBX⁺07] comporte 1907 images scannées de feuilles appartenant à 32 espèces différentes. Un aperçu global des images de cette base est donné par la figure 4.7. Chaque classe (espèce) contient entre 50 et 77 images. Le tableau 4.10 montre la répartition des images dans les classes. Une classe contient en moyenne 59 images. Les espèces sont donc suffisamment représentées. Les images proviennent de scans de feuilles et le fond est complètement uniforme. Les images de cette base sont de bonne qualité. De plus, Flavia ne contient que des feuilles simples et le pétiole est absent sur toutes les images. En fait, les pétioles des feuilles peuvent avoir différentes longueurs et orientations et leur présence peut entraîner une variabilité intra-classe considérable (cf. Figure 4.6). De plus, le pétiole n'est pas utilisé par les botanistes pour identifier l'espèce d'une feuille.

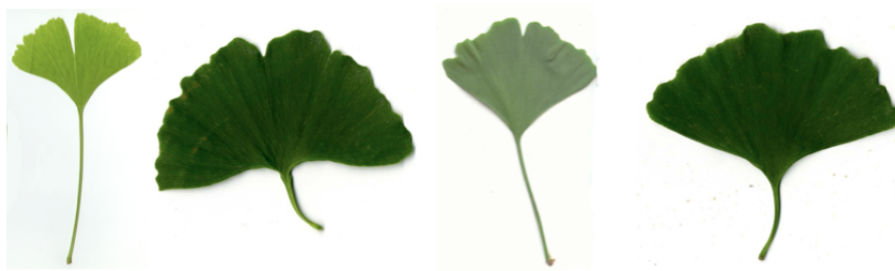


FIGURE 4.6 – Différentes longueurs et orientations du pétiole (exemple de l'espèce Ginkgo).

Cependant, la base Flavia constitue un grand défi à cause de la similarité inter-classe et la variabilité intra-classe. (cf. Figure 4.7).

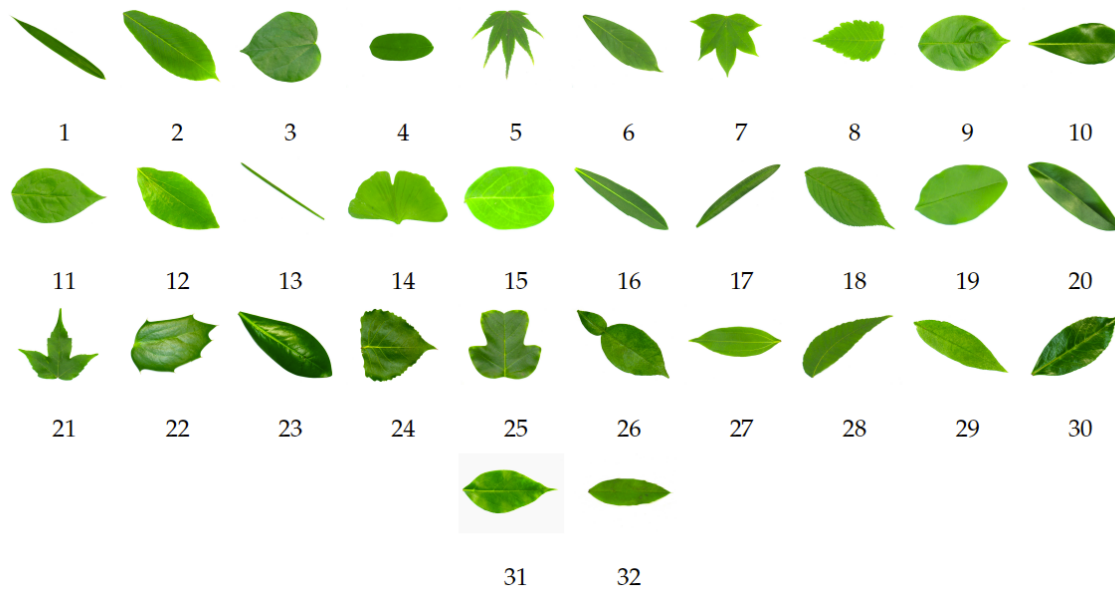


FIGURE 4.7 – Aperçu de la base Flavia. Une image par espèce est montrée.

Espèces (#1 à #16)	Nombre d'images	Espèces (#17 à #32)	Nombre d'images
#1	59	#17	60
#2	63	#18	55
#3	72	#19	55
#4	73	#20	52
#5	56	#21	53
#6	62	#22	55
#7	52	#23	58
#8	59	#24	63
#9	50	#25	53
#10	63	#26	56
#11	52	#27	55
#12	56	#28	65
#13	77	#29	54
#14	62	#30	60
#15	61	#31	65
#16	66	#32	65

TABLE 4.10 – Répartition des images dans la base Flavia

Protocole d'évaluation

Les méthodes évaluées sur Flavia dans la littérature sont essentiellement des approches d'identification d'espèces [KNSS11b, LKS⁺12, KSL14, UU12]. Le protocole d'évaluation suivi par la plupart des méthodes est le calcul de la **MAP (Mean Average Precision)** et des courbes de rappel/précision [LKS⁺12, KSL14]. Le taux d'identification est mesuré dans

$N_s = 10$					
Descripteur	$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 199$
<i>TSL</i>	0.513	0.600	0.659	0.643	0.438
<i>TSLA</i>	0.548	0.656	0.699	0.665	0.468
<i>TOA</i>	0.538	0.645	0.683	0.663	0.512
<i>TAR</i>	0.413	0.483	0.508	0.416	0.342
$N_s = 5$					
Descripteur	$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 199$
<i>TSL</i>	0.486	0.569	0.629	0.631	
<i>TSLA</i>	0.538	0.621	0.651	0.639	0.421
<i>TOA</i>	0.542	0.63	0.649	0.641	0.447
<i>TAR</i>	0.4	0.45	0.46	0.375	0.338

TABLE 4.11 – Variation de n et N_s

[KNSS11b] sur les k images les plus similaires avec $k = 1$, $k = 3$ et $k = 5$. Dans [UU12], un apprentissage est effectué sur 35 images de chaque classe et l'ensemble de test comporte seulement 5 images.

4.2.2 Résultats des descripteurs de forme

Étude de paramètres

Sur cette base, nous souhaitons utiliser les mêmes paramètres pour toutes les représentations triangulaires afin d'effectuer une étude comparative. D'après les expérimentations menées sur les bases Swedish et Flavia, l'utilisation de 400 points pour décrire le contour conduit généralement à un bon résultat. Pour décrire un point p du contour, nous avons étudié différentes représentations en faisant varier n . Nous partons d'une représentation locale, c'est à dire avec $n < \frac{N}{2} - 1$ points du contour voisins de p pour la construction des triangles, ou bien une représentation globale qui prend en compte tout le contour (Pour $N = 400$, une description globale du contour correspond à $n = 199$). Pour décrire chaque portion du contour définie par n nous avons utilisé deux configurations : $N_s = 5$ et $N_s = 10$. Les résultats sont reportés dans le tableau 4.11.

Une autre expérimentation est menée afin de vérifier si l'utilisation d'une distance logarithmique d_2^{log} améliore les résultats du tableau 4.11. Les paramètres utilisés sur la base Swedish sont testés sur Flavia (cf. Tableau 4.12). Pour cette expérimentation, les meilleures valeurs de MAP pour les quatres descripteurs correspondent à $n = 64$ ($N_s = 6$). Cependant, ces résultats sont moins bons que ceux dans le tableau 4.11 obtenus avec $n = 50$, $N_s = 10$ et d_5^{lin} .

d	d_2^{lin}				d_2^{log}			
n	16	32	64	128	16	32	64	128
N_s	8	16	32	64	4	5	6	7
TSL	0.478	0.563	0.605	0.610	0.430	0.558	0.624	0.608
$TSLA$	0.619	0.619	0.651	0.659	0.474	0.617	0.681	0.658
TOA	0.607	0.607	0.641	0.642	0.493	0.592	0.664	0.623
TAR	0.422	0.468	0.486	0.497	0.391	0.479	0.503	0.482

TABLE 4.12 – MAP sur Flavia avec d_2^{lin} et d_2^{log}

Comparaison des descriptions triangulaires

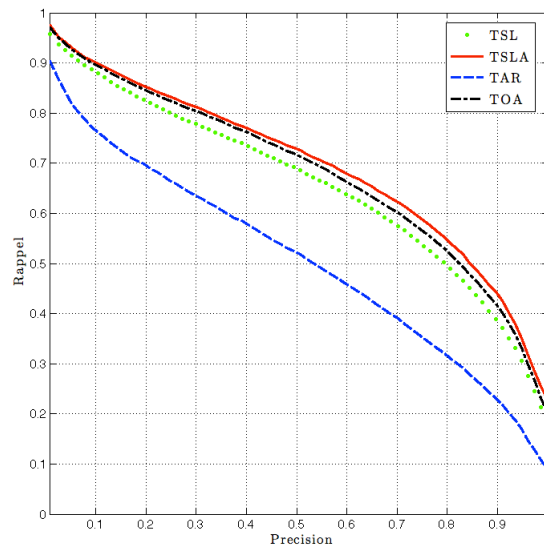


FIGURE 4.8 – Courbes de rappel/précision des descripteurs triangulaires sur la base Flavia

Comparons les courbes de rappel/précision pour les quatre descripteurs TAR , TSL , $TSLA$ et TOA présentées dans la figure 4.8 et étudions les résultats de MAP par classe (cf. Figure 4.9).

La figure 4.8 montre que les trois descriptions proposées TSL , $TSLA$ et TOA sont meilleures que TAR . La meilleure courbe de rappel/précision est obtenue par le descripteur $TSLA$ suivi par TOA et TSL . Ceci donne un aperçu général des performances des descripteurs triangulaires sur la base Flavia. Analysons maintenant les résultats par classe (cf. Figure 4.9).

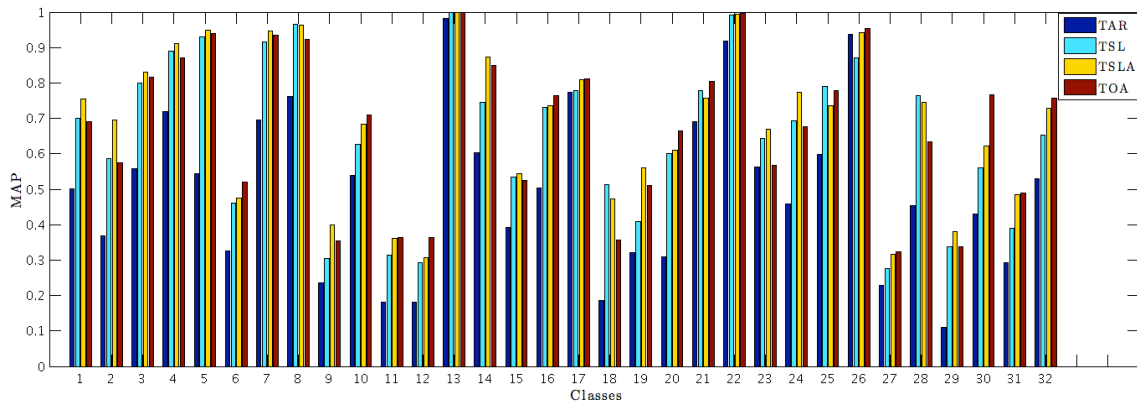


FIGURE 4.9 – Comparaison des descripteurs triangulaires par classe sur la base Flavia

Un premier aspect à analyser, est d'étudier l'apport de l'ajout de l'angle au sommet dans la description *TSLA* en comparant les performances de *TSL* et de *TSLA*. En moyenne sur toutes les classes, *TSLA* améliore la MAP de *TSL* de $+ \cong 3,74\%$. Nous avons observé que l'apport est net sur les feuilles de forme convexe. C'est le cas des classes #2, #14 et #19 où l'amélioration de *TSLA* par rapport à *TSL* est respectivement de $+ \cong 11\%$, $+ \cong 13\%$ et $+ \cong 15\%$.

L'observation de la figure 4.9 permet de voir qu'il existe un écart important de performance en fonction des classes. Par exemple, dans le cas des classes #9, #11, #12, #27 et #29, la valeur de MAP est relativement faible alors que pour les classes #13 et #22, cette valeur est très proche de 1. Examinons les images de ces classes pour pouvoir apporter une interprétation : les classes #9, #11, #12, #27 et #29 présentent une forte similarité inter-espèce : les formes des feuilles de ces espèces sont très proches. (voir Figure 4.10).



FIGURE 4.10 – Similarité de forme entre les classes (de gauche à droite) : #9, #11, #12, #27 et #29

Pour ces classes, *TSLA* et *TOA* obtiennent les meilleures valeurs moyennes de MAP (0.35 et 0.34).

Comparons maintenant les descriptions basées sur les distances ou/et les angles (*TSL*, *TSLA* et *TOA*) et la description qui propose d'utiliser l'aire des triangles (*TAR*). Le premier type est meilleur sur 31 classes parmi 32. Cela confirme que la forme locale d'un triangle est un élément important dans la description du contour. Le seul cas où *TAR* dépasse *TSL* est celui de la classe #26. La particularité des feuilles de cette classe est que

leur contour présente une concavité très caractéristique de l'espèce qui la différencie de toutes les autres classes (cf. Figure 4.11). La description *TAR* propose une bonne approximation des concavités/convexités via les aires signées des triangles qui s'avère bien utile pour cette classe. D'ailleurs, *TOA*, dont les angles orientés approximent aussi les lieux de concavité/convexité, est le premier pour cette classe.

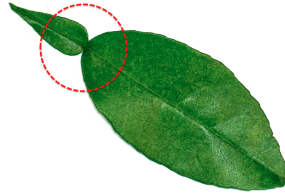


FIGURE 4.11 – concavité caractéristique de l'espèce #26 de la base Flavia

Cas d'occultation partielle

Certaines feuilles sont détériorées ou déchirées, ce qui pourrait correspondre à une occultation partielle. En absence des données nécessaires, nous avons procédé à des bissections manuelles sur les images des feuilles pour simuler leur occultation. Une sélection aléatoire de cinq images de feuille d'espèces différentes de la base Flavia est effectuée. Nous simulons ensuite trois types d'occultations :

1. occultation d'un lobe (Image I_1 dans la figure 4.12),
2. occultation de la moitié de la feuille (Image I_2 , I_3 et I_4 dans la figure 4.12),
3. occultations partielles multiples (Image I_5 dans la figure 4.12).

Les images I_1 , I_2 , I_3 , I_4 et I_5 appartiennent respectivement aux classes #21, #14, #25, #22, et #5 qui présentent différentes formes de feuilles. Le nombre d'images constituant ces classes est indiqué dans le tableau 4.10.

Les occultations manuelles éliminent entre 20% et 50% des points du contour. Les images des feuilles occultées sont présentées dans la figure 4.12. Comme certains points de contour sont plus sensibles (ou inversement plus robustes) à l'occultation partielle, le troisième type d'occultation consiste à simuler des bissections affectant de multiples zones du contour (voir l'image I_5 dans la figure 4.12).

Nous avons effectué des requêtes à partir des images occultées et nous avons calculé la précision sur les quinze premières images retournées. Les mêmes paramètres retenus ($N = 400$, $n = 50$ et $N_s = 10$) sont utilisés dans cette expérimentation. Les résultats sont

présentés dans la table 4.13.

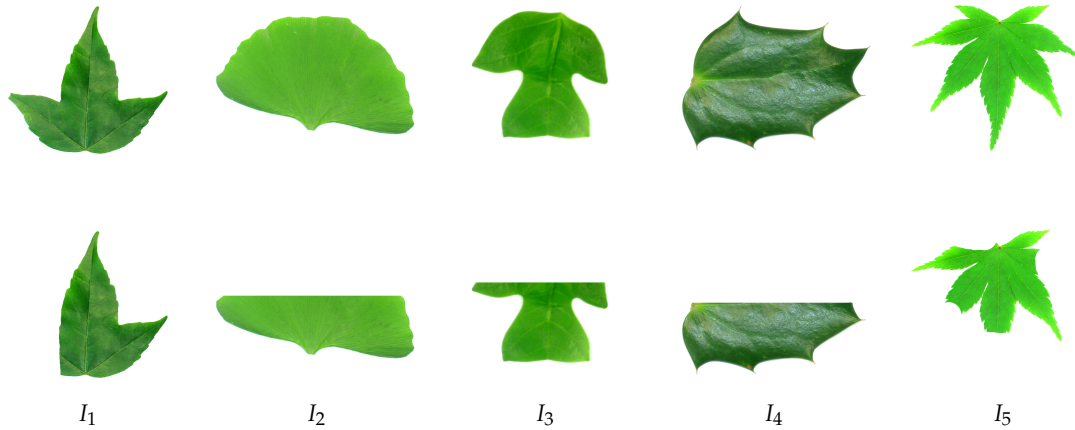


FIGURE 4.12 – Simulation des occultations partielles. Première ligne : images originales. Deuxième ligne : images occultées

La moyenne des taux de classification sur les cinq images est supérieure à 90% pour les descripteurs *TSL*, *TSLA* et *TOA*. Le descripteur *TOA* obtient le meilleur score moyen (96%) sur les images occultées. Il est suivi de *TSL* (94.4%).

Quand les images originales non occultées sont utilisées comme des images requêtes, tous les descripteurs obtiennent 100% de classification correcte considérant les quinze premières images retournées. L'information perdue est mesurée en terme de différence du score d'identification en passant de l'image originale à l'image occultée comme image requête. La quantité d'information perdue avec les images occultées reste faible avec un pourcentage inférieur à 10%. La figure 4.13 montre deux requêtes avec l'image originale et l'image occultée I_5 .

Nos trois descripteurs s'avèrent en général robustes à l'occultation partielle. De plus, notre choix de "matching" local montre son efficacité dans ce cas puisque, même si la donnée d'une partie des points de contour est fausse, les points restants peuvent jouer un rôle important dans la phase de comparaison.

	I_1	I_2	I_3	I_4	I_5	Moyenne	Information perdue
TSL	100	86.6	93.3	100	93.3	94.4	5.4
TSLA	100	100	93.3	100	60	90.6	9.4
TOA	100	100	100	100	80	96	4
TAR	93.3	86.6	80	100	73.3	86.6	13.4

TABLE 4.13 – Scores d'identification avec les images occultées.

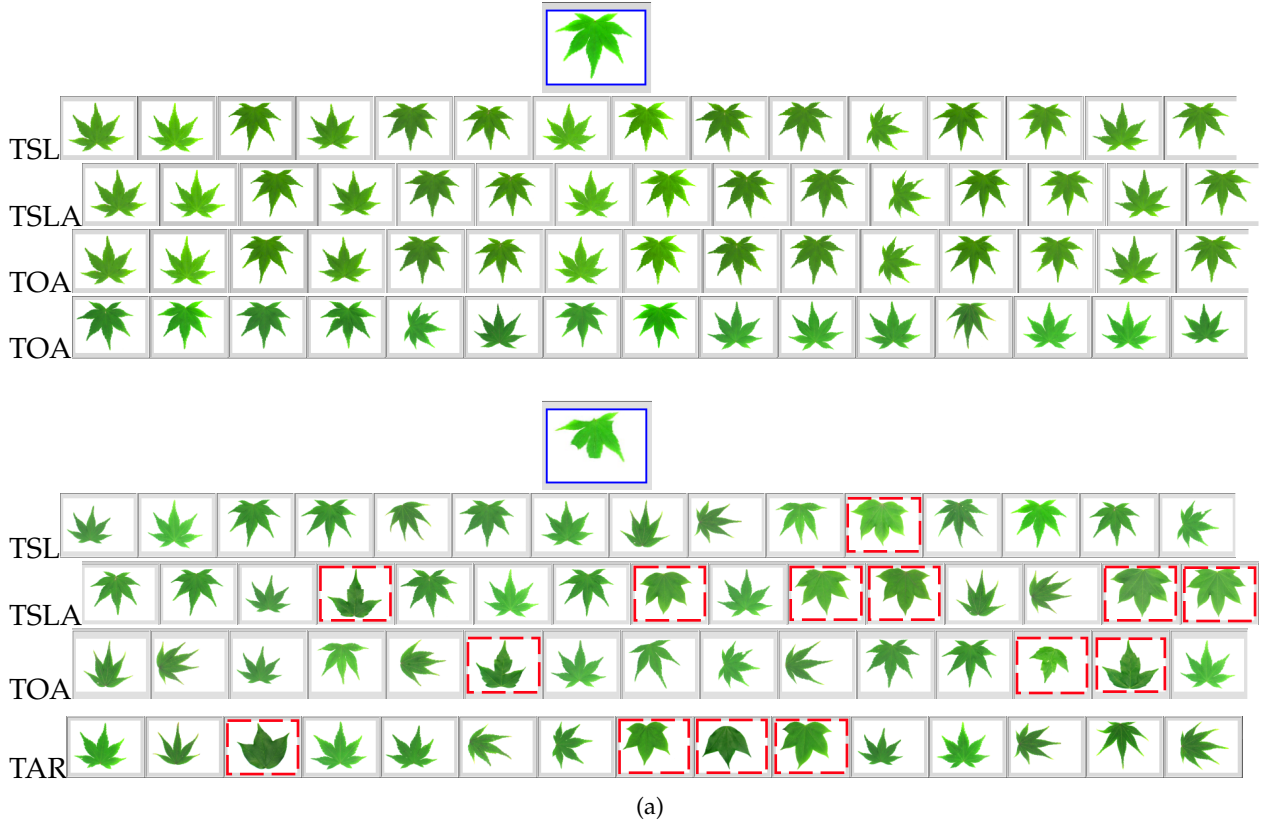


FIGURE 4.13 – Requêtes avec l'image I_5 avant et après occultation. Les faux positifs sont entourés d'une ligne rouge discontinue.

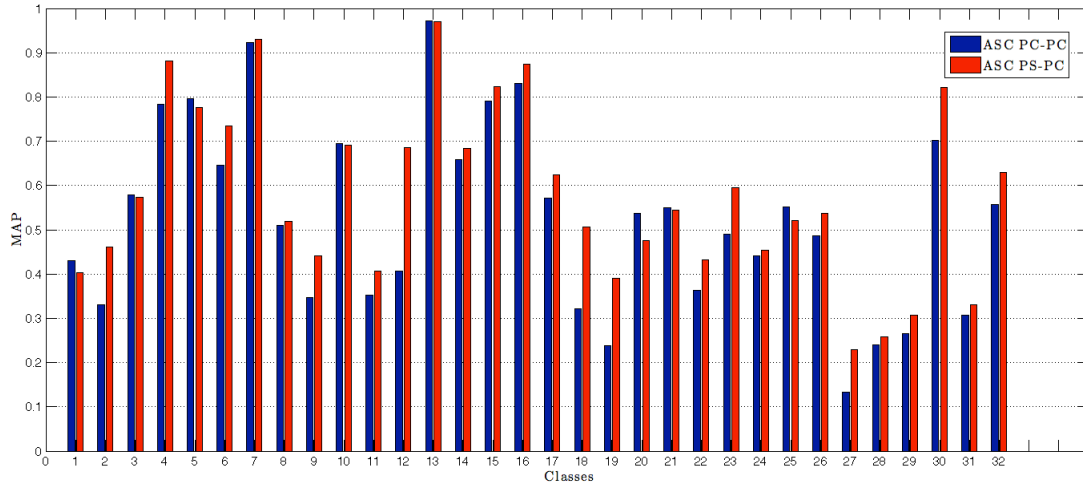
4.2.3 Résultats du contexte de forme étendu (ASC)

Le tableau 4.14 montre les résultats des scénarios du contexte de forme étendu sur Flavia en terme de **MAP** et de **précision**. Nous avons fait varier le nombre de points de l'ensemble de calcul dans chaque scénario.

N		50	100	200	400
ASC_{PC-PC}	Précision	90.14	89.45	90.50	90.81
	MAP	0.537	0.534	0.532	0.529

N_{ps}		50	100	200	400
ASC_{PS-PS}	Précision	55.06	74.20	82.43	87.04
	MAP	0.207	0.326	0.401	0.441
ASC_{PS-PC}	Précision	94.54	96.22	95.85	93.27
	MAP	0.563	0.587	0.586	0.561
$eASC_{PS-PC}$	Précision	86.15	90.61	94.96	96.27
	MAP	0.553	0.624	0.681	0.712

TABLE 4.14 – Résultats des scénarios du contexte de forme étendu utilisant les points saillants

FIGURE 4.14 – Comparaison entre ASC_{PC-PC} et ASC_{PS-PC} par classe sur la base Flavia

Commençons par comparer les deux scénarios ASC_{PC-PC} et ASC_{PS-PC} .

La figure 4.14 présente les valeurs de MAP par classe des descripteurs ASC_{PC-PC} et ASC_{PS-PC} en utilisant respectivement 100 points du contour et 100 points Harris. Une observation générale de l'histogramme permet de constater que ASC_{PS-PC} est généralement meilleur que ASC_{PC-PC} . Sur certaines classes (#3, #8, #10 et #21), la performance des deux descripteurs est presque équivalente.

Analysons les cas où ASC_{PS-PC} se comporte mieux que ASC_{PC-PC} . Ceci est vérifié notamment sur les classes #12 et #18 où la différence de MAP est importante mais également sur les classes #2, #11, #19, #23 et #30. Un aperçu des feuilles de ces classes est donné dans la figure 4.15 avec 100 points de Harris détectés sur chaque feuille. Il est clair qu'il s'agit d'une similarité inter-classe. En effet, la forme des feuilles de ces espèces est similaire. Une description grossière du contour s'avère insuffisante pour les différencier. L'utilisation des points saillants de la feuille dans ce cas est plus pertinente. En effet, nous observons que pour cinq des sept feuilles de forme similaire, une partie des points de Harris sont localisés à l'intérieur du limbe, ce qui réduit considérablement la confusion entre les feuilles de classes différentes car les caractéristiques internes du limbe (nervation, texture, etc.) sont différentes.

En moyenne sur les classes citées ci-dessus, ASC_{PS-PC} améliore la valeur de MAP de ASC_{PC-PC} de $+\cong 14\%$ et la précision de $+\cong 5\%$.



FIGURE 4.15 – Points de Harris détectés sur des feuilles de formes similaires

D'après les résultats du tableau 4.14, nous pouvons noter que :

- Pour ASC_{PS-PS}

Une fois de plus, les résultats de ce scénario sont assez faibles comparés à ceux des autres scénarios. Nous remarquons que sur la base Flavia, la performance de ce scénario s'améliore lorsque le nombre de points saillants augmente mais reste loin d'atteindre une précision très élevée. Cette évaluation confirme les résultats obtenus sur Swedish et nous pouvons conclure que les relations spatiales entre les points saillants ne sont pas déterminantes dans le processus d'identification.

- Pour ASC_{PS-PC}

Lorsque les points saillants sont décrits dans le contexte défini par le contour, les résultats sont assez bons surtout en terme de précision (96.22% pour 100 points saillants utilisés). Un ensemble réduit de points Harris donne généralement de bons scores d'identification. De plus, il existe un gain important en temps CPU par rapport à l'utilisation d'un plus grand nombre de points saillants. En outre, quand le nombre de points saillants augmente les points saillants détectés par le détecteur de Harris ne sont plus caractéristiques de la feuille. Sur Flavia, l'utilisation de 100 points présente un bon compromis.

- Pour $eASC_{PS-PC}$

L'ajout des descripteurs locaux autour des points saillants augmente la valeur de MAP. Par contre, le nombre de points saillants utilisé doit être suffisamment élevé pour que l'apport soit effectif. En effet, pour 50 et 100 points, $eASC_{PS-PC}$ fait baisser la précision de ASC_{PS-PC} en terme de précision. En revanche, à partir de 200 points, La précision et la valeur de MAP sont nettement améliorés. Un autre point qui peut expliquer la bonne performance de ce scénario est que la base Flavia, comme nous l'avons évoqué précédemment, contient des images dont la qualité est assez bonne. Cela contribue à la qualité de la micro-texture de

la feuille. Par conséquent, deux feuilles d’une même espèce auront des caractéristiques visuelles locales proches. Rappelons que le scénario $eASC_{PS-PC}$ n’avait pas amélioré les résultats de ASC_{PS-PC} sur la base Swedish. Nous pouvons conclure que ce scénario est particulièrement intéressant sur les images de feuilles de très bonne qualité.

4.2.4 Résultats des descripteurs de la nervation

Flavia est la seule base sur laquelle tous les descripteurs proposés sont évalués. En effet, cette base contient des images de bonne qualité où les nervures sont assez contrastées.

Étude de paramètres

Échantillonnage des points de nervures

Les réseaux de nervures extraits des images de la base Flavia contiennent généralement un grand nombre de points (En moyenne 34790 points par image). Un échantillonnage de l’ensemble des points de nervure permet de réduire considérablement le temps de calcul. Cinq échantillonnages aléatoires ont été testés, qui gardent respectivement 2%, 4%, 6%, 8% et 10% du nombre total de points de nervures. Nous avons remarqué que les performances enregistrées étaient très proches, avec un gain faible en temps de calcul entre le premier (2%) et le dernier (10%) échantillonnage. Le taux d’échantillonnage a donc été fixé à 10% car il présente le meilleur compromis. Nous n’avons pas testé d’échantillonnage supérieur à 10% pour assurer un temps de calcul raisonnable.

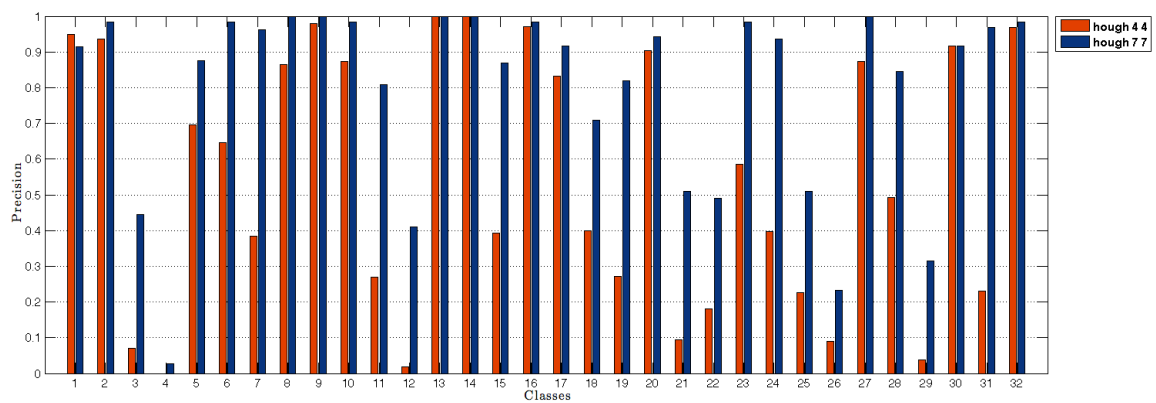


FIGURE 4.16 – Comparaison par classe des deux configurations hough-4-4 et hough-7-7 sur la base Flavia.

Transformée de Hough

La transformée de Hough nécessite la discrétisation de l'espace de projection (A, B) . Nous avons testé deux configurations :

- hough-4-4 : A et B sont discrétisés chacun en 4 partitions.
- hough-7-7 : A et B sont discrétisés chacun en 7 partitions.

Les résultats de ces deux configurations sont présentés dans la Figure 4.16. La configuration hough-7-7 montre une meilleure précision sur 31 des 32 classes de la base Flavia. De plus, la différence entre hough-4-4 et hough-7-7 est importante sur certaines classes (#3, #11, #19, #24, #31). Le temps de calcul du descripteur hough-7-7 est plus long, mais considérant la grande amélioration apportée et la taille raisonnable de la base Flavia, nous avons choisi la configuration hough-7-7 dans la suite de nos expérimentations.

Pour évaluer notre approche sur la base publique Flavia, nous comparons dans un premier temps les résultats obtenus par les descriptions de *Hough*, ASC_{PN-PC} , ASC_{PN-PC} dans le paragraphe 4.2.4. Une comparaison avec les approches existantes est ensuite réalisée dans le paragraphe 4.2.5. Nous nous intéressons dans le paragraphe 4.2.4 à la comparaison des résultats des descriptions ASC_{PN-PN} , ASC_{PN-PC} par rapport à ceux de ASC_{PS-PS} , ASC_{PS-PC} , c'est à dire dans le cas où les points de nervures sont considérés comme les points saillants.

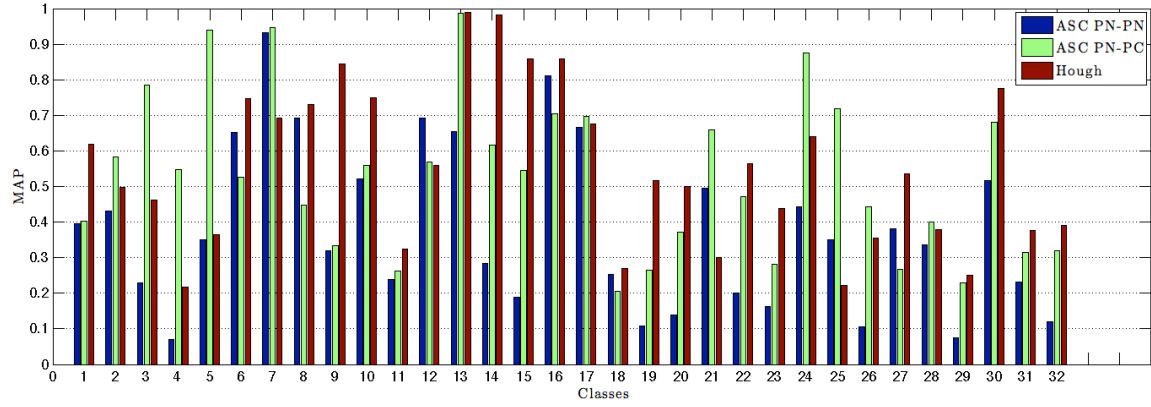
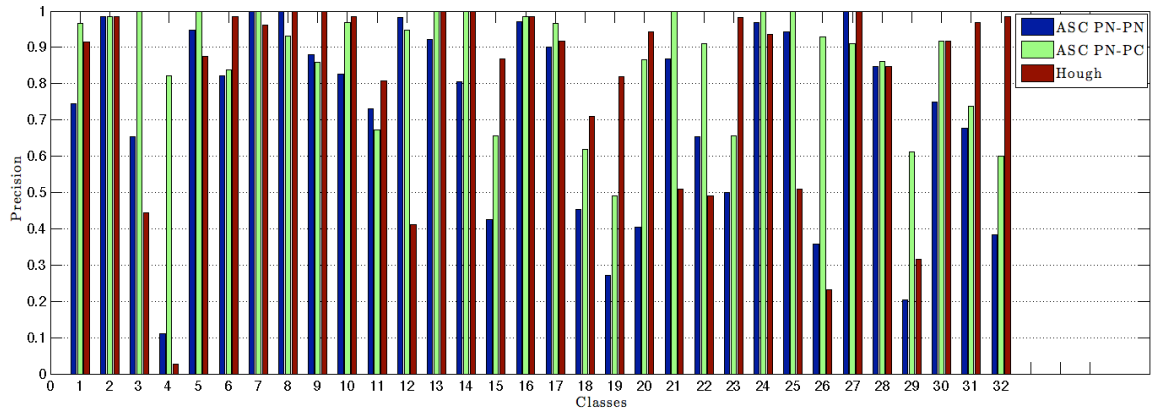
Le protocole d'évaluation utilisé est la valeur de la **MAP (Mean Average Precision)** et la **précision à 1-NN**.

Comparaison des descriptions de nervation Hough, ASC_{PN-PN} et ASC_{PN-PC}

Les figures 4.17 et 4.18 présentent les valeurs de MAP et de précision à $Knn = 1$ pour chaque espèce de la base Flavia des descriptions ASC_{PN-PN} , ASC_{PN-PC} et *Hough*.

Examinons tout d'abord dans la figure 4.17 les valeurs de MAP obtenues par ces trois descriptions. Une première observation de l'histogramme permet de noter que ASC_{PN-PN} est généralement moins performant que les deux autres descriptions. De plus la valeur de MAP de ASC_{PN-PC} est légèrement meilleure que celle de Hough : ASC_{PN-PC} obtient une meilleure MAP pour 21 classes sur 32.

Focalisons-nous sur les classes où ASC_{PN-PC} est nettement supérieur à Hough. C'est le cas des classes #3, #4, #5, #7, #21 et #25.

FIGURE 4.17 – MAP de ASC_{PN-PN} , ASC_{PN-PC} et *Hough* par classe sur la base FlaviaFIGURE 4.18 – Précision de ASC_{PN-PN} , ASC_{PN-PC} et *Hough* par classe sur la base Flavia

Les feuilles de la classe #4 possèdent une nervation particulière qui est caractérisée par une nervure principale souvent très faiblement contrastée. La faible performance du descripteur est due à la qualité du réseau de nervures extrait. En effet, quand les nervures sont peu contrastées, la méthode d'extraction capte essentiellement du bruit (cf. Figure 4.19).

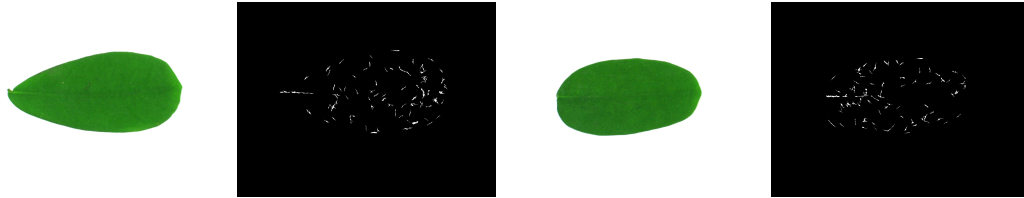


FIGURE 4.19 – Difficulté d'extraction de nervures pour la classe #4 de la base Flavia

Ce phénomène a moins d'influence sur le score du descripteur ASC_{PN-PC} grâce au vote du contour dans la description spatiale.

Les classes #3, #5, #7, et #21 ont en commun une nervation palmée et par conséquent l'arrangement de leurs nervures est proche. Il s'agit d'un cas de similarité inter-classes. Dans la figure 4.20 les signatures visuelles de Hough sur les réseaux de nervures de feuilles issues des classes #5, #7 et #21 sont présentées. Nous remarquons la grande similarité des réseaux de nervures extraits et de leurs descriptions par la transformée de Hough. Par contre, la forme des feuilles est très différente, ce qui explique les bons résultats obtenus par le descripteur ASC_{PN-PC} pour ces espèces.

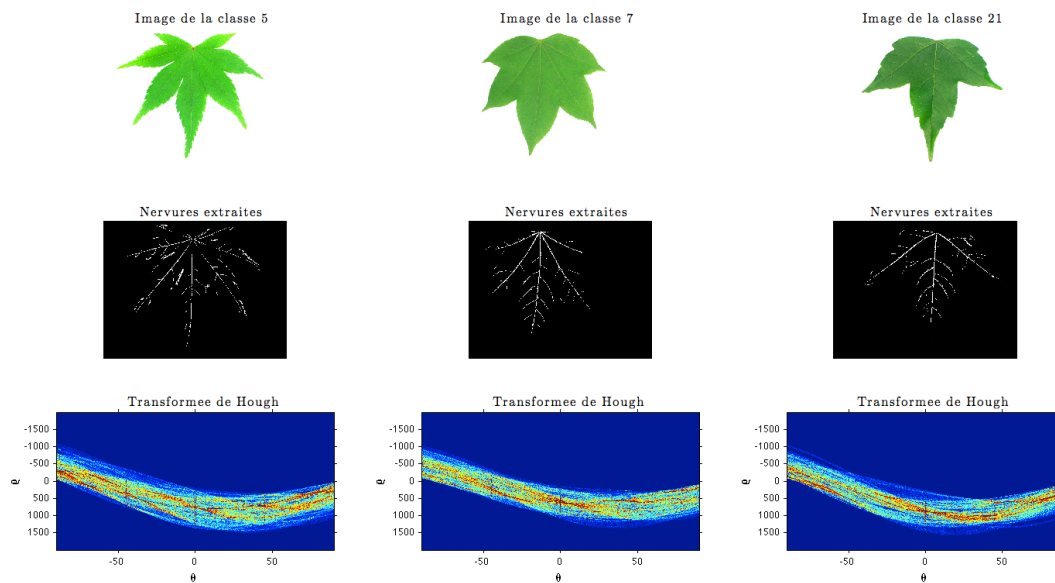


FIGURE 4.20 – Similarité de la nervation des classes #5, #7 et #21. Première ligne : une image couleur de chaque classe. Deuxième ligne : nervures associées. Troisième ligne : transformées de Hough.

Etudions maintenant le cas où la description de Hough est la meilleure. Ceci est visible clairement pour les classes #1, #8, #9 et #14.

Les classes #1 et #14 possèdent respectivement une nervation dichotomique et une nervation parallèle. Ce sont deux types caractéristiques de l'espèce et très différents de celles des autres classes. La Figure 4.21 montre des résultats d'extraction de nervures pour les classes #1 et #14. La structure très particulière des réseaux de nervures rend leur approximation par des droites via la méthode de Hough plus efficace pour les différencier que leur description spatiale. Ceci explique la précision élevée obtenue dans ces cas par la description de Hough : 94% et 100%. Ces résultats sont comparables à ceux obtenus par un descripteur de forme, à l'instar du scénario ASC_{PC-PC} qui obtient une précision de 98% et 100% sur ces deux classes avec 200 points échantillonnés sur le contour.

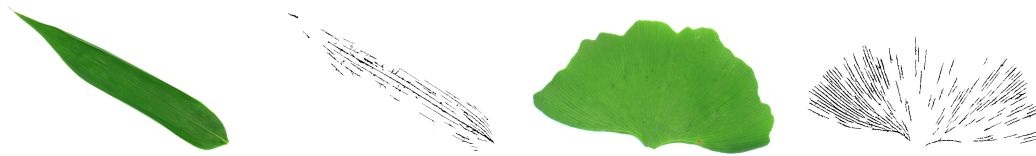


FIGURE 4.21 – Exemples de nervures extraites de feuilles des classes #1 et #14 de la base Flavia.

Les feuilles des classes #8 et #9 ont des formes similaires mais des réseaux de nervures visuellement différents. Comme le montre la Figure 4.22, l'information fournie par le réseau de nervures s'avère plus pertinente que celle du contour. C'est pour cette raison que la description de Hough a de meilleures performances que la description spatiale ASC_{PN-PC} où les points du contour forment l'ensemble votant.

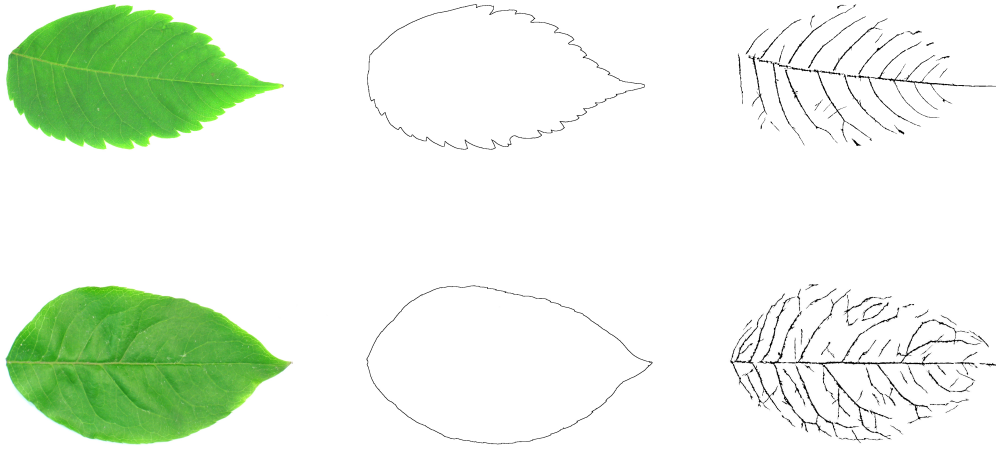


FIGURE 4.22 – Similarité de forme entre les classes #8 et #9 de la base Flavia. Première colonne : images originales. Deuxième colonne : contours. Troisième colonnes : nervures

Comparaison de ASC_{PS-PC} et de ASC_{PN-PC}

Nous avons présenté dans le chapitre précédent le descripteur ASC_{PS-PC} où l'ensemble de calcul \mathcal{C} est composé de points saillants donnés par le détecteur de Harris. Nous nous intéressons dans ce paragraphe à l'évaluation du contexte de forme étendu suivant les deux schémas suivants :

- L'ensemble de calcul \mathcal{C} est composé des points de Harris (ASC_{PS-PC})
- L'ensemble de calcul \mathcal{C} est composé des points de nervures (ASC_{PN-PC}).

ASC	Précision 1-NN (%)	MAP (%)
$ASC_{PS_{50}-PC}$	94.54	56.36
$ASC_{PS_{100}-PC}$	96.22	58.70
$ASC_{PS_{200}-PC}$	95.85	58.69
ASC_{PN-PC}	86.83	53.81

TABLE 4.15 – Comparaison de la performance de ASC_{PS-PC} et de ASC_{PN-PC} .

L'examen des valeurs de précision et de MAP montre une supériorité du scénario (ASC_{PS-PC}). La description de la feuille par ses points saillants s'avère plus efficace que sa description par les points de nervures. En effet, les points saillants représentent des points particuliers en provenance de différentes primitives de la feuille : contour, nervures, texture, etc. Dans le cas des feuilles dentées par exemple, les points de Harris sont surtout localisés sur le contour. ASC_{PS-PC} obtient une meilleure valeur de MAP que ASC_{PN-PC} pour 20

classes sur 32 et dépasse largement ASC_{PN-PC} pour les classes #4, #18, et #32. Les feuilles des classes #4 et #32 ne contiennent pas de nervures apparentes. Ces deux classes sont les plus difficiles pour l'extraction de nervures. Les points de Harris fournissent alors une information plus pertinente car ils sont positionnés sur les éléments saillants du contour ou sur des éléments de texture, qui ont plus d'importance dans le processus d'identification dans le cas de ces classes.

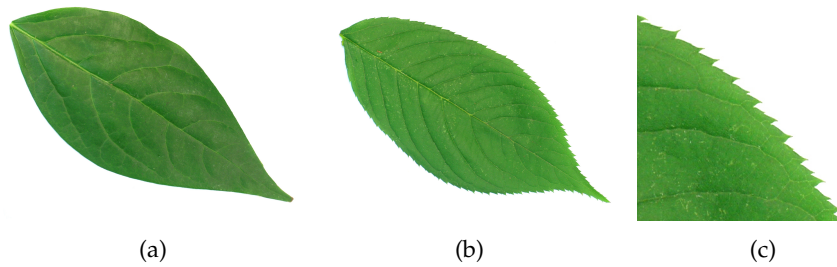


FIGURE 4.23 – Similarité entre les feuilles des espèces #11 et #18. (a) Une feuille de la classe #11. (b) Une feuille de la classe #18. (c) Dents sur le contour caractérisant la classe #18.

Les feuilles de la classe #18 possèdent le même type de nervation que celles de la classe #11 (voir Figure 4.23). Elles ont également des formes très similaires. La différence principale entre les feuilles de ces deux classes se situe au niveau des dents présentes sur le contour des feuilles de la classe #18 (cf. Figure 4.23(c)). Cette information n'est pas prise en compte par ASC_{PN-PC} alors qu'elle est très bien représentée avec les points de Harris dans ASC_{PS-PC} . En effet, les dents sur le contour sont représentées par une suite de coins alternés, efficacement repérés avec le détecteur de Harris (cf. Figure 4.24). Ceci permet d'expliquer la supériorité de ASC_{PS-PC} vis à vis de ASC_{PN-PC} pour la classe #18.



FIGURE 4.24 – Points de Harris sur une portion du contour d'une feuille de la classe #18 de la base Flavia.



FIGURE 4.25 – Exemples de nervures extraites de feuilles des classes #3, #24 et #25.

Le cas où ASC_{PN-PC} obtient une performance largement supérieure que ASC_{PS-PC} est celui des classes #3, #24 et #25. Pour ces classes, ASC_{PN-PC} a une précision moyenne de 100% à 1-NN (97,3% pour ASC_{PS-PC}). Ce sont des exemples d'espèces où la corrélation spatiale entre les nervures et le contour de la feuille est pertinente et assez discriminante. De plus, la nervation des feuilles de ces classes est extraite avec une bonne précision (cf. Figure 4.25).

4.2.5 Comparaison avec l'état de l'art

Plusieurs méthodes décrites dans [LKS⁺12] ont été testées sur la base Flavia. Afin de comparer les performances de nos descripteurs avec ces méthodes, la même métrique d'évaluation est utilisée. Nous calculons la valeur de **la précision moyenne (MAP : Mean Average Precision)** sur toute la base. Des courbes de rappel/précision sont également établies. Les résultats sont reportés dans le Tableau 4.16. Les descripteurs TSL , $TSLA$ et TOA obtiennent une valeur de MAP supérieure à toutes les approches évaluées dans [LKS⁺12]. Parmi les descripteurs triangulaires, la meilleure MAP est obtenue par $TSLA$, elle dépasse largement la meilleure MAP de l'état de l'art (69.93 contre 57.21)•

Les scénarios $ASC_{PS_{100}-PC}$ et $eASC_{PS_{400}-PC}$ obtiennent des valeurs de MAP supérieures à celles dans [LKS⁺12]. Le descripteur $eASC_{PS_{400}-PC}$ s'avère le meilleur sur la base Flavia avec une MAP de **71.2**. Cela montre la pertinence de ce scénario en particulier quand les images sont de bonne qualité, ce qui est le cas des images de la base Flavia.

					Nos descripteurs	MAP
					TSL	0.659
					TOA	0.683
					$TSLA$	0.699
Méthodes dans [LKS ⁺ 12]	D2	MSDM	GEDT	REM	ASC_{PC-PC}^{50}	0.537
					ASC_{PS-PS}^{400}	0.441
MAP	0.428	0.479	0.480	0.572	$ASC_{PS_{100}-PC}$	0.587
					$eASC_{PS_{400}-PC}$	0.712
					ASC_{PN-PN}	0.378
					ASC_{PN-PC}	0.538
					$Hough$	0.558

TABLE 4.16 – Comparaison des résultats de nos descripteurs avec ceux des méthodes évaluées dans [LKS⁺12]

4.3 ÉVALUATION SUR LA BASE ONE-HUNDRED PLANT SPECIES LEAVES (LEAF100)

4.3.1 Caractéristiques de la base Leaf100

La base "One-hundred plant species leaves", que nous dénommons Leaf100, est une nouvelle base de feuilles mise récemment à la disposition du grand public [MCO13]. Elle comporte 1600 images réparties d'une façon uniforme sur 100 espèces contenant des feuilles simples. Le nombre d'espèces est plus important que celui de Swedish (15 espèces) et de Flavia (32 espèces). Les auteurs ne fournissent que les masques binaires des images segmentées. Par conséquent, cette base ne pourra être utilisée que pour l'évaluation des descripteurs de forme. La figure 4.26 montre un aperçu de cette base. Peu de travaux l'ont utilisé vue sa date de création.

La base Leaf100 se caractérise par une très forte similarité inter-classe (cf. Figure 4.26).

4.3.2 Résultats des descripteurs de forme

Sur la base Leaf100, nous avons calculé la précision de la première image retournée ($knn = 1$) et la valeur de MAP.

D'après l'étude de paramètres sur les autres bases, nous avons gardé 400 points échantillonnés pour décrire le contour. Différentes valeurs de n , N_s et m ont été testées. Les résultats sont reportés dans les tableaux 4.17 et 4.18.

Lorsque $n = 199$, ceci correspond à une description complète du contour.

N_s	5					10				
n	10	20	50	100	199	10	20	50	100	199
<i>TSL</i>	0.463	0.543	0.603	0.559	0.367	0.465	0.569	0.652	0.593	0.419
<i>TSLA</i>	0.518	0.606	0.631	0.564	0.394	0.487	0.63	0.66	0.62	0.454
<i>TOA</i>	0.513	0.592	0.564	0.519	0.395	0.491	0.601	0.663	0.57	0.448
<i>TAR</i>	0.258	0.339	0.404	0.334	0.298	0.26	0.367	0.5	0.385	0.367

TABLE 4.17 – MAP des descripteurs triangulaires de forme sur la base Leaf100

N_s	5					10				
n	10	20	50	100	199	10	20	50	100	199
<i>TSL</i>	65.62	79.93	84.18	81.25	72.00	65.62	79.87	89.06	83.62	74.50
<i>TSLA</i>	73.25	86.00	87.00	81.62	74.12	68.12	86.25	89.12	86.25	78.00
<i>TOA</i>	69.75	84.12	81.00	77.50	71.93	66.43	81.81	88	82.31	74.06
<i>TAR</i>	45.31	60.43	69.50	61.25	59.56	42.31	62.43	78.43	68.75	70.00

TABLE 4.18 – Précision ($knn = 1$) des descripteurs triangulaires de forme sur la base Leaf100

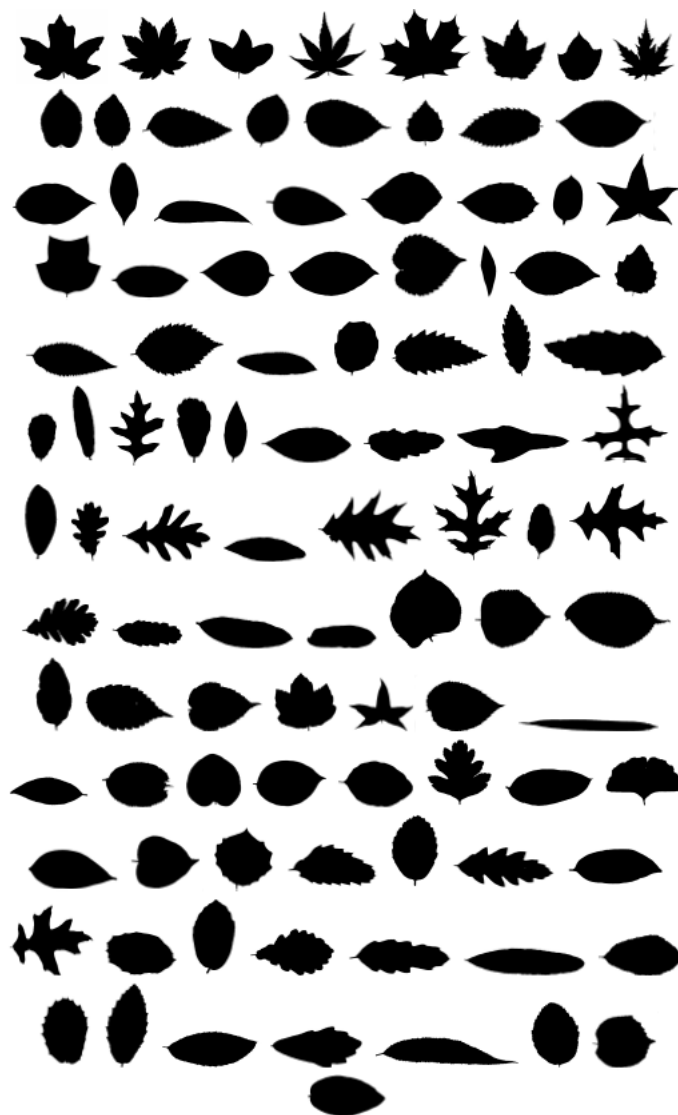


FIGURE 4.26 – Aperçu de la base Leaf100. Une image par classe est montrée

En examinant les résultats, la meilleure configuration pour tous les descripteurs triangulaires correspond aux paramètres suivants :

- $N=400$, $n=5$, d_5^{lin} , $N_s=10$

De plus :

- Les descripteurs *TSL*, *TSL* et *TOA* obtiennent des scores très proches en terme de précision et de MAP. Ils se positionnent tous les trois devant le descripteur *TAR*.
- Les ordres respectifs des résultats de précision et de MAP des descripteurs ne sont pas identiques comme c'est souvent le cas pour les autres bases évaluées. Par exemple, le descripteur *TOA* obtient la troisième meilleure précision alors qu'il a la meilleure valeur de MAP. Ces deux métriques d'évaluation, bien que corrélées, ont une application différente. La précision des k plus proches voisins est généralement utilisée par les systèmes d'identification alors que la MAP est un critère de base dans l'évaluation d'un système de recherche d'images par le contenu visuel.
- En considérant la précision et la valeur de MAP, le descripteur *TSLA* obtient, encore une fois, la meilleure performance sur cette base. Pourtant, l'amélioration apportée par rapport à *TSL* est très faible sur cette base.

Commençons par observer les classes où le descripteur *TAR* dépasse un ou plusieurs autres descripteurs.

En terme de MAP, *TAR* est meilleur que tous les autres descripteurs dans 3% des classes. Il dépasse le descripteur *TSLA* dans 4% des classes. L'écart moyen le plus important est réalisé sur la classe #51 (espèce *Pterocarya-Stenoptera*), cf. Figure) où *TAR* améliore le score moyen des autres descripteur de 15.6%. Quelques images de cette classe sont montrées dans la figure 4.27. Nous remarquons qu'il s'agit de feuilles dentées qui se caractérisent par des variations importantes au niveau du contour. De plus, la forme globale de cette espèce est similaire à celle d'autres classes, notamment des classes *Quercus-Phellos* et *Quercus-Rhysophylla* (cf. Figure 4.28) qui possèdent un contour lisse. Donc, les dents représentent une caractéristique distinctive dans ce cas. Cette information est efficacement décrite par le descripteur *TAR*. Le descripteur *TOA* obtient la deuxième meilleure valeur de MAP sur cette classe.



FIGURE 4.27 – Exemples d'images de l'espèce *Pterocarya-Stenoptera* dans la base Leaf100

Numéro de la classe	Espèce	Précision				MAP			
		<i>TSL</i>	<i>TSLA</i>	<i>TOA</i>	<i>TAR</i>	<i>TSL</i>	<i>TSLA</i>	<i>TOA</i>	<i>TAR</i>
#71	<i>Quercus Kewensis</i>	68.75	81.25	62.50	62.50	0.311	0.388	0.301	0.236
#88	<i>Quercus x Hispanica</i>	43.75	43.75	37.50	31.25	0.269	0.307	0.251	0.184

TABLE 4.19 – Résultats des descripteurs triangulaires sur les deux classes #71 et #88 de la base Leaf100



FIGURE 4.28 – Similarité de forme entre les espèces, de gauche à droite, *Pterocarya-Stenoptera*, *Quercus-Phellos* et *Quercus-Rhysophylla*

Robustesse à la similarité inter-classe

La similarité inter-classe constitue une difficulté classique pour la majorité des descripteurs de forme.

Dans cette expérience, nous avons choisi de considérer le cas des deux espèces *Quercus-x-Hispanica* et *Quercus-Kewensis* qui se caractérisent par une très forte similarité au niveau de la forme, difficilement différentiable (cf. Figure 4.29).

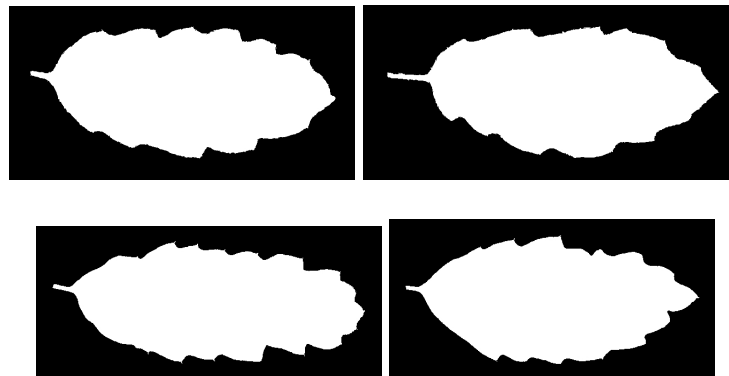


FIGURE 4.29 – Similarité de forme entre les espèces #71 : *Quercus-kewensis* (Première ligne) et #88 : *Quercus-x-Hispanica* (Deuxième ligne)

Les résultats obtenus sur ces deux classes (cf. Tableau 4.29) montrent que le descripteur *TSLA* se comporte le mieux en terme de Précision et de MAP. En deuxième position, nous trouvons le descripteur *TSL*. En effet, l'ajout de l'information sur l'angle interne dans *TSLA* apporte plus de précision que *TSL* (81.25% contre 68.75% pour la première image retournée) dans le cas de l'espèce *Quercus Kewensis* et réduit donc la confusion avec des formes similaires provenant d'autres classes. Cette expérimentation montre que *TSLA* est le descripteur triangulaire le plus approprié pour décrire des formes similaires de la base

Leaf100. Cette observation a été confirmée dans le cas de la base Swedish, précédemment utilisée.

4.4 ÉVALUATION SUR LES SCANS ET LES PSEUDO-SCANS DE FEUILLES D'IMAGECLEF 2011

4.4.1 Caractéristiques de la base ImageCLEF 2011 (Pl@ntLeaves)

ImageCLEF [GBJ⁺11] est une campagne d'évaluation qui comporte plusieurs tâches dont la tâche d'identification des plantes qui nous intéresse. ImageCLEF 2011 propose 3 catégories d'images de feuilles appartenant à 71 espèces :

- des feuilles scannées (scans). Cette catégorie contient des images avec un fond complètement uniforme.
- Pseudo-scans de feuilles. L'acquisition des images est réalisée avec un appareil de photographie sur un fond uniforme. Les pseudo-scans se caractérisent par un bruit dû au protocole d'acquisition comme par exemple les ombres portées, le contraste de la feuille, luminance irrégulière, etc. Cette catégorie contient également des feuilles mortes.
- Des photographies de feuilles avec un fond naturel.

Chaque catégorie est subdivisée en une base d'apprentissage et une base de test. Le but consiste donc à trouver l'espèce de chaque feuille de test. Le score d'identification S proposé par ImageCLEF est différent de celui calculé par des métriques classiques. Ce choix est guidé par deux hypothèses :

- Les feuilles d'un même arbre (i.e d'un même individu) sont plus similaires entre elles que deux feuilles de la même espèce mais provenant de deux arbres différents.
- Les photos prises par une même personne vont avoir un protocole d'acquisition similaire, et donc peuvent comporter des similarités visuelles indépendantes des feuilles scannées ou photographiées.

Nous nous intéressons ici aux deux premières catégories d'images (scans et pseudo-scans). (cf. Figure 4.30). Le tableau 4.20 montre la répartition des images dans les deux ensembles d'apprentissage et de test.

Le nombre d'individus désigne le nombre d'arbres à partir desquels les images sont prises. Le nombre de contributeurs désigne le nombre de personnes participant à l'acquisition

Catégorie	Ensemble	Nombre d'images	Nombre d'individus	Nombre contributeurs
Scans	ensemble d'apprentissage	2349	151	17
	ensemble de test	721	55	13
Pseudo-scans	ensemble d'apprentissage	717	51	2
	ensemble de test	180	13	1

TABLE 4.20 – Composition des ensembles d'apprentissage et de test de la base Pl@ntLeaves dataset



FIGURE 4.30 – Scans et pseudo-scans de la base ImageCLEF 2011. une image par espèce est montrée. Trois premières lignes : exemples de scans. Deux dernières lignes : exemples de pseudo-scans

d'images. Ces deux paramètres sont importants dans le calcul du score d'identification S décrit dans le paragraphe suivant.

Protocole d'évaluation

Il s'agit de trouver l'espèce de chaque image de test en effectuant une recherche d'images similaires dans la base d'apprentissage. Le score d'identification S est un score de classification normalisée par espèce, individu (arbre) et contributeur . La formulation mathématique de S est comme suit :

$$S = \frac{1}{U} \sum_{u=1}^U \frac{1}{P_u} \sum_{p=1}^{P_u} \frac{1}{N_{u,p}} \sum_{p=1}^{N_{u,p}} s_{u,p,n} \quad (4.2)$$

U : nombre de contributeurs. Un contributeur est une personne qui a contribué avec au moins une image de feuille se trouvant dans la base de test.

P_u : nombre d'individus observés par le $u^{\text{ème}}$ contributeur.

$N_{u,p}$: nombre d'images prises du $p^{\text{ème}}$ individu observé par le $u^{\text{ème}}$ contributeur.

$s_{u,p,n}$: Score de classification (1 or 0) pour la $n^{\text{ème}}$ image prise sur $p^{\text{ème}}$ individu par le $u^{\text{ème}}$ contributeur.

4.4.2 Résultats des descripteurs de forme sur ImageCLEF2011

$N_s = 10$										
Catégorie	Scans					Pseudo-scans				
n	10	20	50	100	199	10	20	50	100	199
TSL	0.581	0.692	0.802	0.768	0.688	0.554	0.621	0.757	0.705	0.618
$TSLA$	0.577	0.699	0.796	0.751	0.676	0.565	0.646	0.780	0.724	0.621
TOA	0.593	0.702	0.794	0.769	0.689	0.569	0.657	0.779	0.726	0.637
TAR	0.522	0.643	0.721	0.711	0.622	0.497	0.549	0.636	0.615	0.532

TABLE 4.21 – Scores d'identification sur les scans et les pseudo-scans de test d'ImageCLEF2011 en faisant varier n

Pour les deux catégories (scan et pseudo-scan), les meilleurs paramètres sont :

- $N = 400$, $n = 50$, d_5^{lin} . Cela correspond à $N_s = 10$.

D'après les scores d'identification reportés dans le tableau 4.21, l'utilisation d'une valeur de $n \ll \frac{N-1}{2}$ conduit généralement à une meilleure performance qu'une description globale ($n = 199$). Ceci est vérifié quand $n \geq 20$. Sur les scans, la performance des descripteurs triangulaires TSL , $TSLA$ et TOA est très proche. Sur les pseudo-scans, les meilleurs scores d'identification de $TSLA$ et TOA sont presque identiques et sont légèrement supérieurs à celui de TSL .

La présence de pétiole sur un certain nombre d'images de test rend l'identification par la forme plus difficile. En effet, les feuilles de même espèce peuvent avoir des pétioles de forme très différente (longueur, orientation, etc.) comme le montre la figure 4.31. En outre, la présence de bourgeon sur certaines images modifie la forme globale de la feuille (cf. feuille à droite dans la figure 4.31)).

Les images de test peuvent contenir des images de feuilles dont le contour est déformé (cf. Figure 4.32). Ces déformations sont souvent dues à des facteurs externes affectant la



FIGURE 4.31 – Différentes formes de pétioles des scans d’ImageCLEF 2011



FIGURE 4.32 – Déformations de contour d’une feuille de même espèce

feuille (maladies foliaires, perte de folioles sur les feuilles composées, etc.). Certaines déformations de contour peuvent être vues comme des occultations partielles comme dans la figure 4.32. La base ImageCLEF 2011 permet donc d’évaluer la robustesse des descripteurs triangulaires face aux déformations de contour.

4.4.3 Résultats du contexte de forme étendu sur les scans et les pseudo-scans de test

Descripteur	Scans					Pseudo-cans				
	20	50	100	200	400	20	50	100	200	400
ASC_{PC-PC}	0.544	0.702	0.705	0.703	0.717	0.493	0.686	0.677	0.683	0.691
ASC_{PS-PS}	0.115	0.399	0.487	0.533	0.548	0.09	0.291	0.322	0.359	0.366
ASC_{PS-PC}	0.647	0.702	0.669	0.616	0.616	0.6371	0.657	0.638	0.627	0.609
$eASC_{PS-PC}$	0.510	0.603	0.578	0.707	0.598	0.208	0.406	0.377	0.559	0.556

TABLE 4.22 – Scores d’identification du contexte de forme étendu sur la base de feuilles ImageCLEF2011

La base de test d’ImageCLEF 2011 (scans + pseudo-scans) contient un certain nombre d’images de feuilles mortes 4.33. Sur ces feuilles, il est moins évident de repérer des points saillants pertinents. La nervation et la texture sont moins apparentes et difficilement détec-



FIGURE 4.33 – Exemples d’images de feuilles mortes dans les pseudo-scans de test d’ImageCLEF2011

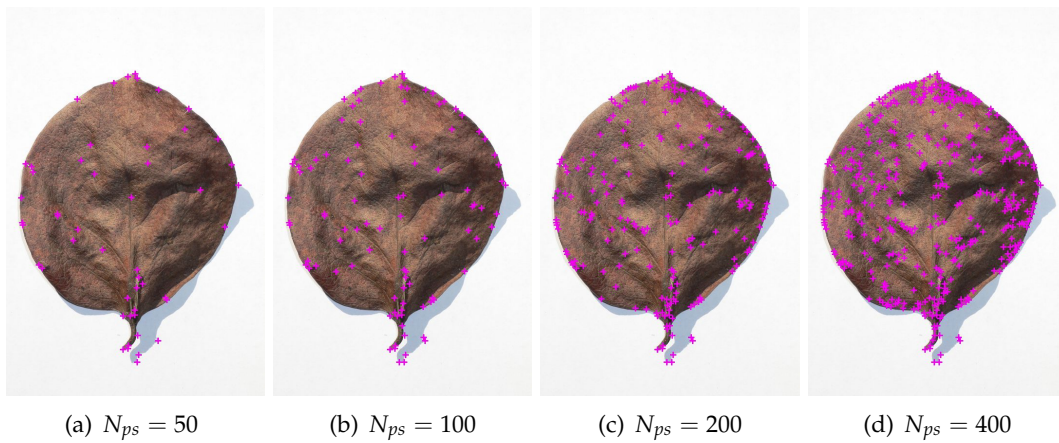


FIGURE 4.34 – Détection de points saillants sur une feuille morte de la base de test d’ImageCLEF2011

tées.

La figure 4.34 montre différents nombres de points de Harris détectés sur une image contenant une feuille morte. Nous remarquons que les points de Harris détectés ne décrivent pas forcément des caractéristiques pertinentes. De plus, un certain nombre de points saillants est localisé sur le fond, en particulier sur les coins de l’ombre de la feuille.

Résultats du contexte de forme étendu sur la base d’apprentissage (scan et pseudo-scan) d’ImageCLEF 2011

Dans cette expérimentation, nous avons fusionné la base d’apprentissage des scans et des pseudo scans d’ImageCLEF2011 en une seule base comportant 3066 images. Nous cherchons ici à évaluer les scénarios du contexte de forme étendu sur un grand nombre d’images. De plus, les scans et les pseudo-scans ont un protocole d’acquisition différent. Cela permet d’évaluer les descripteurs sur des images de différente qualité. Deux ensembles

de points saillants sont utilisés :

- un ensemble qui comporte un nombre réduit de points saillants (50 points).
- un ensemble qui comporte un nombre assez grand de points saillants (400 points).

Cette expérimentation ne vise pas à étudier le nombre de points saillants optimal car cette étude sera faite dans la suite dans l'évaluation des descripteurs sur l'ensemble de test. Le but ici est d'évaluer le contexte de forme étendu sur un grand nombre d'images (3066 images).

La valeur de MAP est calculée pour chaque scénario et des courbes de rappel/précision sont également établies dans la figure 4.35.

Les résultats sont reportés dans le tableau 4.23. Nous dénommons *std* la description qui calcule des descripteurs locaux (voir paragraphe 3.2.3, page 69) autour des points saillants.

Scénario	ASC_{PC-PC}	ASC_{PS-PS}	ASC_{PS-PC}	<i>std</i>	$eASC_{PS-PC}$
MAP (50 points)	0.36	0.14	0.40	0.45	0.50
MAP (400 points)	0.40	0.25	0.42	0.57	0.68

TABLE 4.23 – MAP des différents scénarios du contexte de forme étendu sur la base d'apprentissage (scans+pseudo-scans) d'ImageCLEF 2011

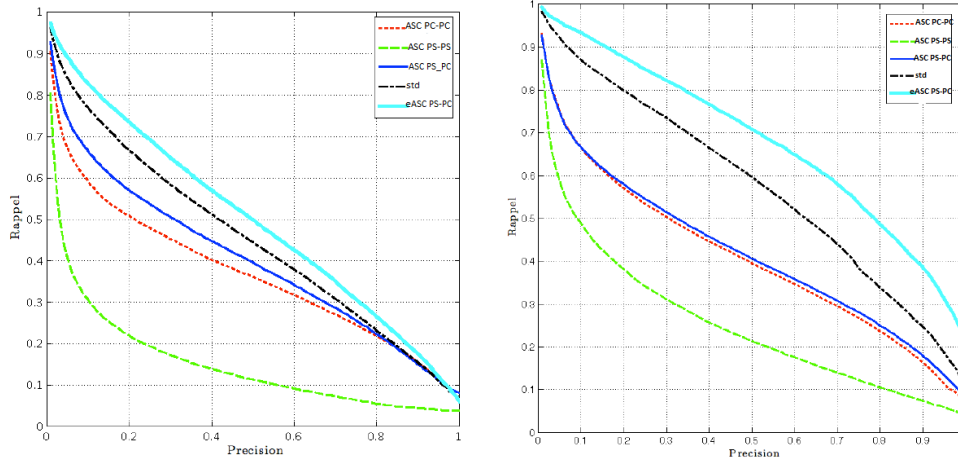


FIGURE 4.35 – Courbe de rappel/précision avec 50 (à gauche) et 400 (à droite) points saillants

Le tableau 4.23 et les courbes de rappel/précision de la figure 4.35 permettent d'observer que :

- tous les descripteurs obtiennent de meilleurs résultats avec 400 points. Cependant, nous remarquons que l'amélioration apportée dans le cas de ASC_{PS-PC} est minime. Les courbes de rappel/précision de ASC_{PS-PC} ont presque la même allure en passant de 50 à 400 points saillants. Cela confirme notre observation faite sur les autres bases

sur le fait qu'à partir d'un certain seuil, le détecteur de Harris commence à détecter des points non pertinents pour l'identification des feuilles. Nous avons remarqué que si l'ensemble de calcul \mathcal{C} est de taille réduite dans la description ASC_{PS-PC} , cela réduit considérablement le temps de calcul sans pour autant conduire à une baisse importante de performance.

- le scénario ASC_{PS-PS} obtient le moins bon résultat. La même conclusion faite lors des évaluations sur les autres bases est confirmée. En effet, les relations spatiales entre les points de Harris détectées ne sont pas assez discriminantes pour l'identification. La corrélation spatiale entre les points Harris et le contour de la feuille est beaucoup plus pertinente.
- ASC_{PS-PC} donne une meilleure courbe de rappel/précision que de ASC_{PC-PC} . L'écart est surtout clair quand l'ensemble de calcul est de taille réduite (50 points). Avec 400 points, les résultats des deux descripteurs sont très proches. Un grand apport de ASC_{PS-PC} vis à vis de ASC_{PC-PC} est qu'il est capable d'atteindre des résultats meilleurs ou comparables avec un nombre de points saillants beaucoup plus petit. La dimension de l'ensemble de calcul joue un rôle important, surtout dans l'étape de mise en correspondance entre les images. En effet, si nous supposons que nous allons à titre d'exemple faire une recherche exhaustive sur notre base, chaque point de l'ensemble de calcul de taille n_{ec} , p_i $i = \{1..n_{ec}\}$ est comparé à $(n \times 3065)$ points. Le nombre total de comparaisons effectuées est : $(n_{ec} \times 3065 \times (3066))$. Le coût de comparaison est extrêmement plus important quand n_{ec} passe de 50 à 400.
- Les descripteurs locaux *std* associés à des points Harris obtiennent une bonne valeur de MAP qui s'améliore considérablement lorsque 400 points Harris sont utilisés. Cela montre la pertinence des descriptions locales dans le voisinage des points saillants. En revanche, cela nécessite un coût considérable en temps de calcul surtout lors de la phase de comparaison et de mise en correspondance discutée ci-dessus.
- le descripteur *eASC* (enriched ASC) réalise le meilleur résultat en terme de MAP et de courbe rappel/précision. La combinaison de la description spatiale (ASC_{PS-PC}) et de la description locale (*std*) s'avère très utile pour caractériser une feuille sur la base considérée.

Afin de mieux illustrer l'apport du contexte de forme étendu via le scénario ASC_{PS-PC} par rapport à la description classique contexte de forme ASC_{PC-PC} , nous comparons dans la Figure 4.36 les réponses de deux requêtes avec les descripteurs ASC_{PC-PC} et ASC_{PS-PC} sur

une feuille *Rhamnus alaternus*. Cette espèce a une forme globale similaire à celle d'autres espèces dans la base. Nous cherchons à travers cette expérimentation à étudier le comportement des descripteurs ASC_{PC-PC} et ASC_{PS-PC} dans le cas de similarité inter-classe. Nous analysons les résultats sur les quinze premières images retournées ($knn=15$).

Avec le descripteur ASC_{PC-PC} , 4 images sur 15 sont pertinentes, c'est à dire de la même espèce que l'image requête. Ceci donne un taux de précision de 27%. Cela illustre les limites du descripteur ASC_{PC-PC} et des descriptions grossières de la forme en général, et confirme qu'une information supplémentaire est nécessaire pour être suffisamment robuste à la similarité inter-espèces.

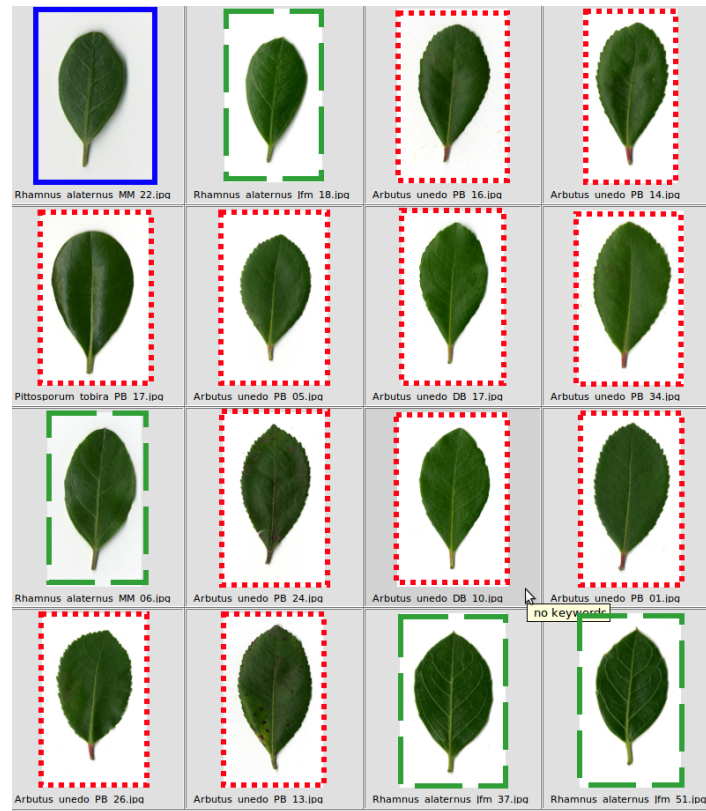
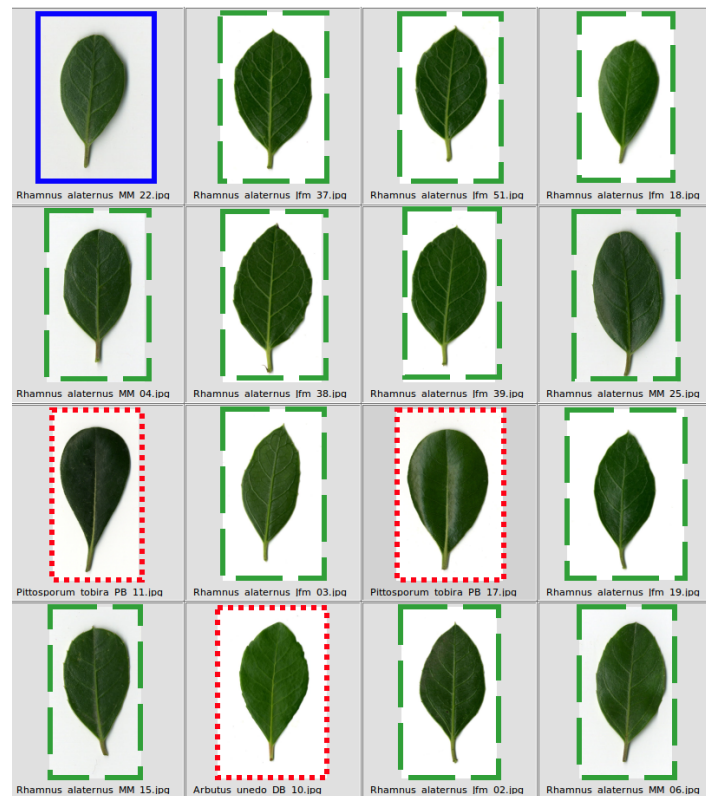
La figure 4.36(b) montre le résultat de la même requête en utilisant notre descripteur ASC_{PS-PC} . Le taux de précision passe à 80%. Cela s'explique par le fait que ASC_{PS-PC} inclut une information importante sur la corrélation spatiale entre les points saillants et le contour. L'autre avantage de ASC_{PS-PC} est le fait que l'information contour n'est pas ignorée : elle forme l'ensemble votant dans ASC_{PS-PC} . De plus, les points Harris peuvent être localisés sur le contour, en particulier pour les feuilles lobées et dentées.

La figure 4.37 montre l'ensemble de points similaires lors de la phase de mise en correspondance. Dans le cas de ASC_{PC-PC} , un certain nombre de points de l'image requête est considéré comme similaire à celui de l'image retournée alors que l'image retournée appartient à une autre espèce. Cela est dû à la grande similarité de la forme globale des feuilles des deux espèces.

Dans le cas, de ASC_{PS-PC} les points saillants détectés sont répartis sur le contour, la nervation et à l'intérieur du limbe. L'image retournée est de même espèce que l'image requête. Notons ici que nous avons utilisé ASC_{PS-PC} eu lieu de $eASC_{PS-PC}$ pour prouver que même en l'absence des descriptions locales std , la corrélation spatiale entre les points saillants et le contour peut jouer un rôle primordial dans l'identification.

4.4.4 Comparaison avec l'état de l'art

Le tableau 4.24 présente les scores d'identification sur la base ImageCLEF 2011. La majorité des descripteurs proposés dans cette thèse obtiennent des scores meilleurs que ceux des participantes d'ImageCLEF 2011. Les descripteurs triangulaires se comportent le mieux sur cette base même par rapport à DFH+GP [YMB12] qui obtient de très bons scores d'identification sur les scans et les pseudo-scans.

(a) ASC_{PC-PC} (b) ASC_{PS-PC} FIGURE 4.36 – Deux réponses avec la même image requête utilisant ASC_{PC-PC} et ASC_{PS-PC} .

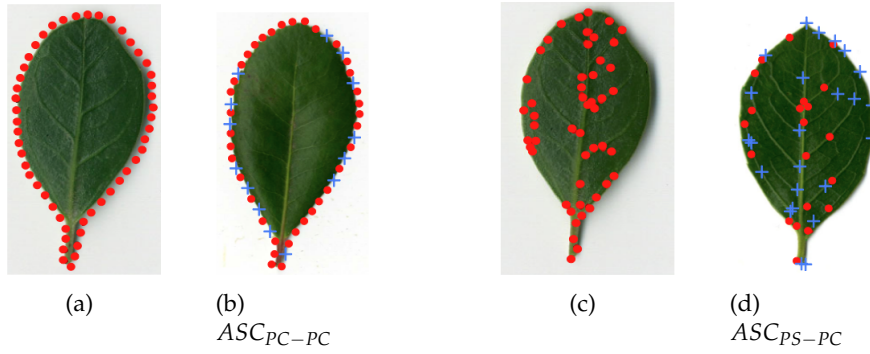


FIGURE 4.37 – (Mise en correspondances (matching) en utilisant ASC_{PC-PC} et ASC_{PS-PC} . (a) image requête **Rhamnus alaternus** MM de la figure 4.36 et points impliqués dans le calcul de ASC_{PC-PC} (b) image retournée en utilisant ASC_{PC-PC} : les points rouges sont les points mis en correspondance (c) même image requête et points saillants détectés (d) image retournée en utilisant ASC_{PS-PC} et points mis en correspondance (en rouge)

Participants	Score d'identification	
	Scans	Pseudo-scans
IFSC USP run2	0.562	0.402
inria imedia plantnet run1	0.685	0.464
IFSC USP run1	0.411	0.430
LIRIS run3	0.546	0.513
LIRIS run1	0.539	0.543
Sabancı-okan-run1	0.682	0.476
LIRIS run2	0.530	0.508
LIRIS run4	0.537	0.538
inria imedia plantnet run2	0.477	0.554
IFSC USP run3	0.356	0.187
DFH+GP [YMB12]	0.778	0.725
<i>TSL</i>	0.802	0.757
<i>TOA</i>	0.794	0.780
<i>TSLA</i>	0.796	0.779
<i>TAR</i>	0.721	0.636
ASC_{PC-PC}	0.717	0.691
ASC_{PS-PS}	0.548	0.366
ASC_{PS-PC}	0.702	0.669
$eASC_{PS-PC}$	0.707	0.559

TABLE 4.24 – Comparaison avec les résultats de l'état de l'art sur la base ImageCLEF 2011

4.5 EVALUATION SUR LES SCANS ET LES PSEUDO-SCANS DE LA BASE IMAGE-CLEF 2012

ImageCLEF 2012 est composée des images de 2011 et elle est enrichie par de nombreuses nouvelles images et de nouvelles espèces [GBJ⁺12]. Elle contient 126 espèces.

Les scans de feuilles sont répartis en 4870 images pour l'apprentissage et 1760 images de test. Pour la catégorie pseudo-scan, l'ensemble d'apprentissage comporte 1819 images et l'ensemble de test est composé de 907 images. Le nombre d'images et le nombre de contributeurs dans les deux catégories sont présentés dans le tableau 4.25.

La métrique d'évaluation est légèrement modifiée par rapport à ImageCLEF 2011 mais le score S désigne toujours la même classification normalisée. La différence réside dans la manière de calculer $s_{u,p,n}$ dans l'équation 4.2 (page 132) : le rang de l'espèce correcte de chaque image de test est calculé dans la liste des espèces retournées. Un score $\frac{1}{k}$ est attribué à chaque image de test avec k le rang de l'espèce correcte : 1 si la première espèce retournée correspond à celle de l'image de test.

Catégorie	Ensemble d'images	Nombre d'images	Nombre de contributeurs
Scans	ensemble d'apprentissage	4870	22
	ensemble de test	1760	10
Pseudo-scans	ensemble d'apprentissage	1819	8
	ensemble de test	907	10

TABLE 4.25 – Composition des ensembles d'apprentissage et de test de la base d'images de feuilles d'ImageCLEF2012

4.5.1 Résultats des descripteurs de forme

Nous faisons varier les paramètres n et N_s d'une façon similaire à celle dont nous avons procédé sur les autres bases (voir Tableau 4.26). La meilleure valeur n correspond approximativement au quart du contour ($n=50$ pour $N=400$). Encore une fois, la description locale s'avère plus performante que la description globale. Les meilleurs paramètres sont similaires à ceux trouvés sur la majorité des bases, à savoir :

- $N = 400$, $n = 50$ et d_5^{lin} ($N_s = 10$).

La performance des descripteurs triangulaires est très proche sur les scans et sur les pseudo-scans d'ImageCLEF 2012.

Catégorie	Scan					Pseudo-scan				
n	10	20	50	100	199	10	20	50	100	199
TSL	0.41	0.484	0.52	0.492	0.402	0.413	0.571	0.61	0.595	0.536
$TSLA$	0.457	0.509	0.53	0.519	0.418	0.465	0.588	0.632	0.607	0.569
TOA	0.449	0.512	0.54	0.527	0.433	0.449	0.583	0.631	0.609	0.572
TAR	0.4	0.451	0.52	0.511	0.407	0.388	0.477	0.514	0.505	0.411

TABLE 4.26 – Scores d'identification sur les scans et les pseudo-scans de test d'ImageCLEF2012 en faisant varier n

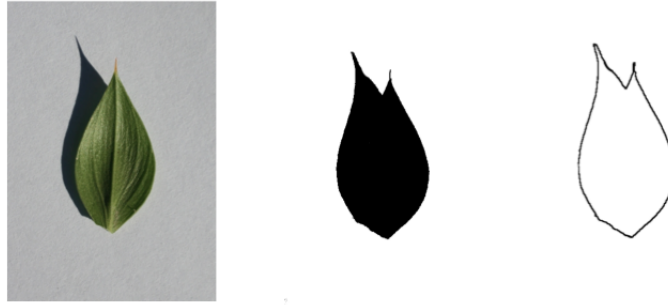


FIGURE 4.38 – Erreur de segmentation sur les pseudo-scans de test d’ImageCLEF2012

Notons la difficulté d’extraction de contour sur certaines images de test, en particulier les pseudo-scans. En effet, l’ombre de la feuille dû au protocole d’acquisition, produit souvent des erreurs de segmentation (cf. Figure 4.38).

4.5.2 Résultats du contexte de forme étendu

Les scores d’identification des scénarios du contexte de forme étendu sont montrés dans le tableau 4.27. Le descripteur ASC_{PS-PC} se comporte le mieux sur les scans et améliore légèrement les résultats de ASC_{PC-PC} . Sur les pseudo-scans, la description de forme via ASC_{PC-PC} s’avère beaucoup plus efficace que le reste des scénarios utilisant les points saillants. Comme nous l’avons remarqué précédemment, cela est dû principalement aux bruits affectant les images 4.38. D’ailleurs, l’ajout des descripteurs locaux dans $eASC_{PS-PC}$ n’améliore pas les résultats obtenus par ASC_{PS-PC} .

Descripteur	Score sur les scans	Score sur les pseudo-scans
ASC_{PC-PC}^{50}	0.465	0.581
ASC_{PC-PC}^{100}	0.507	0.583
ASC_{PC-PC}^{200}	0.522	0.586
$ASC_{PS_{50}-PC}$	0.506	0.284
$ASC_{PS_{100}-PC}$	0.510	0.409
$ASC_{PS_{200}-PC}$	0.482	0.397
$eASC_{PS_{50}-PC}$	0.492	0.191
$eASC_{PS_{100}-PC}$	0.463	0.326
$eASC_{PS_{200}-PC}$	0.387	0.299

TABLE 4.27 – Scores d’identification du contexte de forme étendu sur les scans et les pseudo-scans de ImageCLEF2012

Participants	Type de méthode	Score d'identification	
		Scans	Pseudo-scans
SABANCI OKAN run 2	semi-automatique	0.58	0.55
THEWHO run 3	semi-automatique	0.43	0.4
SABANCI OKAN run 1	automatique	0.58	0.55
IFSC USP run 1	semi-automatique	0.35	0.41
LIRIS Reves run 1	semi-automatique	0.42	0.51
INRIA Imedia PlantNet run 1	automatique	0.49	0.54
INRIA Imedia PlantNet run 2	automatique	0.39	0.59
THEWHO run 4	semi-automatique	0.37	0.35
LSIS DYNi run 3	automatique	0.41	0.42
THEWHO run 1	semi-automatique	0.37	0.34
<i>TSL</i>	automatique	0.52	0.61
<i>TOA</i>	automatique	0.54	0.63
<i>TSLA</i>	automatique	0.53	0.63
<i>TAR</i>	automatique	0.52	0.51
ASC_{PC-PC}^{200}	automatique	0.52	0.59
$ASC_{PS_{100}-PC}$	automatique	0.51	0.41
$eASC_{PS_{50}-PC}$	automatique	0.49	0.33

TABLE 4.28 – Scores d'identification sur les scans et les pseudo-scans de la base ImageCLEF 2012

4.5.3 Comparaison avec l'état de l'art

Le tableau 4.28 montre les scores d'identification obtenus par nos différents descripteurs en comparaison avec ceux obtenus par les méthodes évaluées sur les scans et les pseudo-scans de la base ImageCLEF2012 [GBJ⁺12]. Nous avons participé à cette campagne d'évaluation avec une méthode d'identification des scans et une méthode sur les pseudo-scans [BYM⁺12]. L'identifiant de notre participation est "INRIA Imedia PlantNet run 2". Nous avons utilisé la méthode ASC_{PC-PC}^{200} + mise en correspondance locale sur les pseudo-scans. Elle a obtenu le meilleur score d'identification parmi les méthodes participantes. Sur les scans, la méthode $eASC_{PS-PC}^{200}$ + mise en correspondance locale a été sélectionnée car elle a donné le meilleur score d'identification sur la base d'apprentissage mais ceci n'a pas été vérifié sur les scans de la base de test à cause de la qualité des images de test. Les descripteurs triangulaires obtiennent des scores d'identification très semblables. Sur la catégorie des pseudo-scans, *TSLA* et *TOA* obtiennent le meilleur score et dépassent toutes les autres approches. Sur les scans de feuille, les descripteurs triangulaires obtiennent le deuxième meilleur score d'identification.

4.6 ÉVALUATION SUR LA BASE IMAGECLEF2013

La base ImageCLEF2013 est enrichie par rapport à ImageCLEF 2012. Les scans et les pseudo-scans sont regroupés pour constituer une unique base. Elle comporte 9781 images dans la base d'apprentissage et 1250 dans la base de test. Ces images représentent 250 espèces. La base ImageCLEF2013 est l'une parmi les bases disponibles contenant le plus d'images et d'espèces.

Cela permet d'évaluer les descripteurs proposés sur un volume considérable d'images de feuilles. En effet, quand le nombre d'espèces et d'images augmente, les cas de similarité inter-classe et de variabilité intra-classe sont plus courants et par conséquent permettent d'évaluer la robustesse des méthodes d'identification.

La répartition des images, de contributeurs et du nombre d'individus (arbres ou plantes) dans les bases d'apprentissage et de test est reportée dans le tableau 4.29.

Ensemble	Nombre d'images	Nombre d'individus	Nombre de contributeurs
ensemble d'apprentissage	9781	732	36
ensemble de test	1250	14	150

TABLE 4.29 – Composition des ensembles d'apprentissage et de test de la base d'images de feuilles d'ImageCLEF 2013

4.6.1 Résultats des descripteurs de forme

Pour évaluer les descripteurs triangulaires de forme sur la base de feuilles d'ImageCLEF2013, le même nombre de points échantillonnés sur le contour $N = 400$ est utilisé. Nous avons testé les trois configurations suivantes :

1. $n = 40, d_2^{lin}$: pour décrire un point du contour p_i , une portion représentant 20% du contour est retenue. Elle est décrite par 20 triangles espacés de 2 points ($m = 2$). Le but étant de tenir compte des détails de la portion du contour retenue en choisissant une distance réduite entre deux triangles successifs ($m = 2$).
2. $n = 50, d_5^{lin}$: pour décrire un point du contour p_i , une portion représentant le quart (25%) du contour est retenue. Elle est décrite par 10 triangles espacés de 5 points (d_5^{lin}). Ces paramètres ont montré une certaine stabilité sur les bases précédemment évaluées.
3. $n = 100, d_{10}^{lin}$: pour décrire un point du contour p_i , une portion représentant 50% du contour est retenue. Elle est décrite par 10 triangles espacés de 10 points ($m = 10$). Avec ces paramètres, nous considérons une valeur plus élevée de n (longueur de la portion du



FIGURE 4.39 – Difficulté d'extraction du contour sur la base de test d'ImageCLEF 2013

contour retenue).

n	$n = 40, d_2^{lin}$	$n = 50, d_5^{lin}$	$n = 100, d_{10}^{lin}$
N_s	20	10	10
<i>TSL</i>	0.565	0.606	0.545
<i>TSLA</i>	0.569	0.629	0.551
<i>TOA</i>	0.589	0.602	0.548
<i>TAR</i>	0.506	0.534	0.493

TABLE 4.30 – Scores des descripteurs triangulaires sur la base ImageCLEF2013

D'après les scores d'identifications affichés dans le tableau 4.30, les meilleurs paramètres qui donnent les meilleurs résultats pour les quatre descripteurs sont :

- $N + 400, n = 50, d_5^{lin}, N_s = 10$.

Ces paramètres se sont avérés stables sur l'ensemble des bases évaluées. Les scores obtenus par *TSL*, *TSLA* et *TOA* sont proches. La supériorité de la description *TSLA* se confirme sur cette base avec le meilleur score d'identification (0.629) comparé aux autres descripteurs triangulaires.

Il est à noter aussi la difficulté d'extraction du contour sur certaines images de la base de test d'ImageCLEF 2013 (cf. Figure 4.39). Cela est dû essentiellement aux problèmes liés au protocole d'acquisition (contraste insuffisant, ombres portées, etc.). De plus, le seuillage d'Otsu obtient généralement des résultats très satisfaisants sur les images de feuilles avec un fond complètement uniforme (Scans) mais ce n'est pas toujours le cas sur les pseudo-scans où les erreurs de segmentation sont courantes.

4.6.2 Résultats du contexte de forme étendu

Le tableau 4.31 montre les scores d'identification des scénarios du contexte de forme étendu sur les images de test d'ImageCLEF2013. Le descripteur ASC_{PC-PC} se comporte le

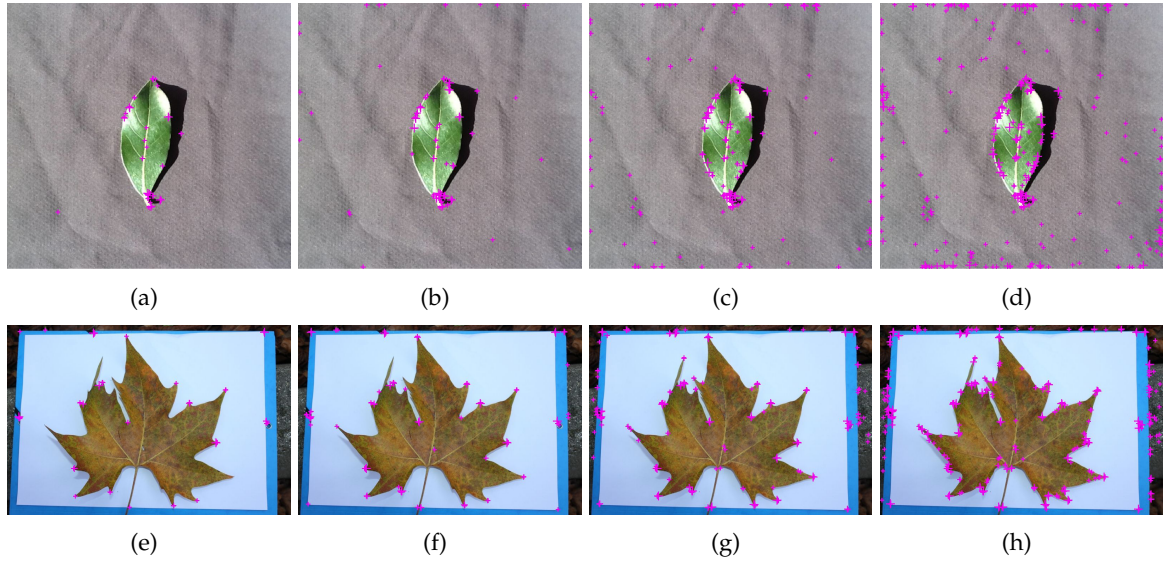


FIGURE 4.40 – Points saillants détectés sur des images de test d'ImageCLEF 2013. De gauche à droite : 50, 100, 200 et 400 points de Harris.

mieux sur la base de test. Le fait de regrouper les scans et les pseudo-scans dans la même base rend un descripteur de forme plus efficace qu'un descripteur de points saillants. Nous avons discuté de la difficulté d'extraction de points saillants pertinents sur les bases de test d'ImageCLEF2011 et d'ImageCLEF2012. La même observation peut être faite sur la base de test d'ImageCLEF 2013. En effet, en plus de certaines erreurs de segmentation, les points de Harris sont localisés parfois sur le fond comme illustré dans la figure 4.40. Cela est dû au bruit de fond caractérisant certaines images.

Descripteur	Score
ASC_{PC-PC}^{50}	0.452
ASC_{PC-PC}^{100}	0.576
ASC_{PC-PC}^{200}	0.581
$ASC_{PS_{50}-PC}$	0.478
$ASC_{PS_{100}-PC}$	0.364
$ASC_{PS_{200}-PC}$	0.273
$eASC_{PS_{50}-PC}$	0.348
$eASC_{PS_{100}-PC}$	0.282
$eASC_{PS_{200}-PC}$	0.251

TABLE 4.31 – Scores d'identification du contexte de forme étendu sur la base de feuilles ImageCLEF2013

4.6.3 Comparaison avec l'état de l'art

Comparons la performance de notre approche utilisant chaque descripteur avec celle des méthodes participantes à la campagne ImageCLEF2013 5.10.

Le descripteur *TSLA* se distingue encore une fois en réalisant la meilleure performance par rapport à toutes les autres méthodes. Les descripteurs triangulaires obtiennent généralement de très bons scores malgré les erreurs de segmentation discutées précédemment : *TSL* et *TOA* sont positionnés en deuxième et troisième position par rapport à la meilleure performance des méthodes participantes.

Il est clair que, sur cette base, les descripteurs de forme, y compris la description ASC_{PC-PC} , obtiennent des scores d'identification meilleurs que tous les scénarios utilisant les points saillants. Le fait de regrouper les scans et les pseudo-scans dans une seule base rend la tâche d'extraction des points saillants pertinents plus difficile.

10 Meilleures Participations	Score d'identification S
Sabanci Okan Run 1	0.607
Inria PlantNet Run 2	0.577
Inria PlantNet Run 3	0.572
Inria PlantNet Run 1	0.557
Inria PlantNet Run 4	0.517
NlabUTokyo Run 1	0.509
NlabUTokyo Run 3	0.502
NlabUTokyo Run 2	0.502
Liris ReVeS Run 2	0.416
Liris ReVeS Run 1	0.412
<i>TSL</i>	0.606
<i>TSLA</i>	0.629
<i>TOA</i>	0.602
<i>TAR</i>	0.534
ASC_{PC-PC}^{200}	0.581
$ASC_{PS_{50}-PC}$	0.478
$eASC_{PS_{50}-PC}$	0.348
Nos descripteurs	

TABLE 4.32 – Comparaison des méthodes évaluées sur la base ImageCLEF 2013

4.7 CONCLUSION

Nous avons évalué les descripteurs proposés sur six bases publiques. Le tableau 4.33 montre les descripteurs évalués sur chaque base et leurs rangs respectifs. Pour le calcul des rangs, le score moyen tenant compte de la valeur de MAP et de la précision est considéré

pour les bases Flavia et Leaf100. Pour les bases ImageClef en 2011 (IC2011) et en 2012 (IC2012), le score moyen des catégories scan et pseudo-scan est utilisé pour le calcul des rangs des descripteurs.

Bases d'images	Swedish	Flavia	Leaf100	IC 2011	IC 2012	IC 2013
Descripteurs de forme						
<i>TSL</i>	3 _(95.73)	4	3	3	3	2
<i>TSLA</i>	1 _(96.53)	2	1	1	2	1
<i>TOA</i>	2 _(96.00)	3	2	2	1	3
Contexte de forme étendu (ASC)						
<i>ASC_{PC-PC}</i>	5 _(91.73)	6	-	4	4	4
<i>ASC_{PS-PS}</i>	7 _(80.05)	7	-	7	-	-
<i>ASC_{PS-PC}</i>	4 _(93.57)	5	-	5	5	5
<i>eASC_{PS-PC}</i>	6 _(91.03)	1	-	6	6	6
Descripteur de nervation						
<i>Hough</i>	-	8	-	-	-	-
<i>ASC_{PN-PN}</i>	-	10	-	-	-	-
<i>ASC_{PN-PC}</i>	-	9	-	-	-	-

TABLE 4.33 – Bases d'images de feuilles utilisées pour l'évaluation des descripteurs

Nous pouvons remarquer :

- la pertinence de la description de la forme foliaire et en particulier des descripteurs triangulaires proposés. Le descripteur *TSLA* a montré une stabilité de performance et il a obtenu le meilleur résultat sur la plupart des bases évaluées. La description *TOA* a montré également une bonne précision, sa performance est comprable à celle de *TSLA*.

L'étude des paramètres menée sur les différentes bases a permis de trouver des paramètres assez stables qui permettent généralement d'obtenir une performance optimale. Ces paramètres sont les suivants :

- $N = 400$, $n = 50$ et d_5^{lin} ($N_s = 10$).

- Nous avons évalué et comparé les différents scénarios du contexte de forme étendu sur cinq bases. Le scénario *ASC_{PS-PC}* qui vise à représenter la corrélation spatiale entre les points saillants et le contour externe d'une feuille s'est avéré utile pour décrire les feuilles. L'avantage principal de ce scénario est le fait d'atteindre une précision d'identification élevée avec un ensemble de calcul de taille relativement réduite. Cela a une importance majeure dans l'étape de mise en correspondance des

images. Lorsque ce scénario est enrichi avec les descripteurs locaux ($eASC_{PS-PC}$), cela n'améliore pas forcément sa performance. Nous avons déduit que cela dépend de la qualité des images de feuilles (feuilles mortes, ombres portées, etc.). En effet, le scénario $eASC_{PS-PC}$ a obtenu la meilleure MAP et la meilleure précision ($knn=1$) sur la base Flavia qui contient des images de feuilles de bonne qualité. En revanche, la performance a baissé sur les bases ImageCLEF en particulier sur la catégorie des pseudo-scans.

- Nous avons évalué trois aspects du réseau de nervures : ses directions principales (*Hough*), les relations spatiales entre les points de nervures (ASC_{PN-PN}) et sa corrélation spatiale avec le contour (ASC_{PN-PC}). Les trois descriptions ont obtenu de bons résultats avec une supériorité de *Hough*.

Dans ce chapitre, tous les descripteurs sont évalués individuellement. Nous nous proposons dans le prochain chapitre d'étudier la fusion des différentes descriptions proposées.

Chapitre 5

Fusion des descriptions foliaires

DANS ce chapitre, nous nous proposons d'étudier les fusions des descripteurs présentés dans les chapitres précédents. Rappelons que nous avons introduit les descripteurs foliaires suivants :

- un descripteur de forme qui inclut trois représentations différentes : *TSL*, *TSLA*, et *TOA*.
- une famille de descripteurs de points saillants via les scénarios du contexte de forme étendu *ASC*.
- trois descriptions de nervation

La fusion utilisée est la fusion tardive, qui est une fusion des rangs établie après la construction des listes des images similaires calculées à partir d'une image requête I et un ensemble d'images \mathcal{D} . Les méthodes de fusion tardive que nous avons utilisées sont présentées dans un premier temps dans la section 5.1.

Nous présentons également le cas des requêtes multiples : c'est à dire celui où une requête visuelle est composée d'un ensemble d'images de feuilles issues d'un même individu (arbre) au lieu d'une image unique.

5.1 MÉTHODES DE FUSION

En examinant les résultats présentés dans [HT05, LW12, RS03], nous avons choisi la fusion basée sur le rang des images dans les listes de similarité.

Supposons que $L_1 \dots L_q$ soient q listes ordonnées contenant les p images les plus similaires à une image requête I , suivant les descripteurs $d_1 \dots d_q$.

Soit $L_m = I_1^m(I) \dots, I_p^m(I)$, $I_l^m(I)$ la l^{me} image de \mathcal{D} la plus similaire à I utilisant le descripteur d_m .

Dans la suite, J appartient à $U = \bigcup_{m=1}^q L_m$, $q > 1$ et $1 \leq m \leq q$. Si $J \in L_m$, nous notons $r_m(J)$ l'index l tel que $J = I_l^m(I)$.

Pour considérer le cas de présence/absence d'images dans les listes L_m lors de la combinaison, nous introduisons les fonctions $h : U \times \{1 \dots q\} \rightarrow \{0, 1\}$ et $w_m : U \rightarrow \{0, \dots, p\}$ définies par :

$$h(J, m) = \begin{cases} 1 & \text{si } J \in L_m \\ 0 & \text{sinon} \end{cases}$$

$$w_m(J) = h(J, m) \times (p + 1 - r_m(J))$$

Nous avons utilisé les méthodes suivantes pour fusionner les listes L_1, \dots, L_q :

A.1 L'algorithme Borda Count (BC) [JHDH06]

$$BC(J) = \sum_{m=1}^q w_m(J)$$

La liste finale contient les images de U triées dans un ordre décroissant suivant leur valeur BC .

A.2 L'algorithme Inverse Rank Position (IRP) [JHDH06]

$$IRP(J) = 1 / \left[\sum_{m \in \{1, \dots, q\} \text{ et } h(J, m) = 1} \frac{1}{r_m(J)} \right]$$

La liste finale est ordonnée suivant un ordre croissant des valeurs IRP

Si les listes $L_1 \dots L_q$ ne contiennent pas le même ensemble d'images, deux méthodes de fusion supplémentaires $BCCombMNZ$ et $IRPCombMNZ$, sont utilisées,

B.1 L'algorithme $BCCombMNZ$ (CombMNZ dans [Lee97]) : où le résultat de fusion est obtenu en multipliant la valeur BC par le nombre de listes qui ont un score h différent de zéro :

$$BCCombMNZ(J) = \left(\sum_{m=1}^q h(J, m) \right) \times BC(J)$$

B.2 L'algorithme $IRPCombMNZ$ étend de la même manière la méthode IRP :

$$IRPCombMNZ(J) = \left(\sum_{m=1}^q h(J, m) \right) \times IRP(J)$$

Comme notre but est l'identification de l'espèce, la fusion peut être effectuée sur LS_m , la liste ordonnée d'espèces établie à partir de la liste L_m des images les plus similaires. Deux méthodes de fusion sont introduites et la liste résultante des espèces candidates est ordonnée suivant leur valeurs de probabilité décroissantes WP ou MP :

C.1 Probabilité pondérée WP :

$$WP(s) = \sum_{m=1}^q W_m(s) \times P_m(s) \quad \text{avec} \\ W_m(s) = \max_S [P_m(s) / (\sum_{m=1}^q \max_S (P_m(s)))]$$

C.2 Probabilité maximale MP :

$$MP(s) = \max_{1 \leq m \leq q} [P_m(s)]$$

5.2 FUSION DE LA DESCRIPTION DE LA FORME ET DES POINTS SAILLANTS

Dans cette section, nous nous intéressons à la fusion de la description triangulaire de forme (TSL, TSLA, TOA) et du descripteur ASC_{PS-PC} sur plusieurs bases. Commençons par la base Flavia.

5.2.1 Comparaison et fusion des descripteurs sur la base Flavia

Dans le cas où la valeur de MAP est calculée, la totalité de la réponse à chaque requête est utilisée et qui est égale au nombre d'images dans la base y compris l'image requête. Les fusions basées sur les probabilités MP et WP ne pourront pas être utilisées dans ce cas là. Il est de même pour les méthodes $BCCombMNZ$ et $IRPCombMNZ$.

Descripteurs	IRP	BC
$TSL + ASC_{PS-50-PC}$	69.87	67.79
$TSLA + ASC_{PS-50-PC}$	71.81	68.96
$TOA + ASC_{PS-50-PC}$	70.46	68.35
$TAR + ASC_{PS-50-PC}$	63.34	62.54
$TSL + ASC_{PS-100-PC}$	70.93	68.11
$TSLA + ASC_{PS-100-PC}$	72.85	69.29
$TOA + ASC_{PS-100-PC}$	71.48	68.73
$TAR + ASC_{PS-100-PC}$	64.48	63.04

TABLE 5.1 – MAP de la fusion des descripteurs triangulaires de forme avec le descripteur de points saillants ASC_{PS-PC}

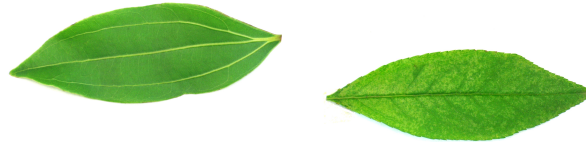
Le tableau 5.7 montre les résultats de fusion des descripteurs de forme et de ASC_{PS-PC} en faisant varier le nombre de points de l'ensemble de calcul.

En moyenne sur les huit configurations, la méthode de fusion IRP est la meilleure. Avec cette méthode, l'apport de la fusion est toujours positif par rapport aux résultats des descripteurs individuels. La meilleure valeur de MAP obtenue résulte de la fusion de $TSLA$ et de $ASC_{PS_{100}-PC}$ avec la méthode IRP .

Retenons ces paramètres et analysons de plus près les résultats dans le cas de similarité inter-classe. Prenons l'exemple des classes #18 et #19 et des classes #27 et #29.



(a) Similarité entre les classes #18 et #19



(b) Similarité entre les classes #27 et #29

FIGURE 5.1 – Similarité inter-classe sur la base Flavia

Nous avons discuté dans les chapitres précédents le cas de similarité de forme inter-classe et son impact sur la performance des descripteurs de forme. Ce cas est encore plus complexe quand il s'agit des données botaniques. Nous avons montré des exemples dans le chapitre 4 où il est très difficile de différencier, même visuellement, les feuilles de différentes espèces.

La solution la plus appropriée consiste alors à inclure d'autres informations sur la feuille provenant d'autres caractères. Dans cette optique nous avons développé trois descripteurs de forme, de points saillants et de nervation dans le but d'augmenter l'efficacité de la description foliaire. Dans cette expérimentation, nous combinons la description de forme avec celle des points saillants.

Le tableau 5.2 montre les résultats obtenus pour les classes (#18, #19) et (#27, #29). Sur toutes ces classes, la fusion améliore les résultats obtenus par les descripteurs individuels en terme de MAP et Précision. Ceci est illustré dans la figure 5.2. Nous remarquons aussi que, généralement, le descripteur ASC_{PS-PC} est plus performant que $TSLA$ sur les classes de formes similaires. Malgré le vote du contour dans le descripteur ASC_{PS-PC} , les histogrammes spatiaux sont calculés réellement sur les points saillants (points de Harris). En effet, dans le cas de formes similaires, ASC_{PS-PC} décrit des caractéristiques différentes (ensembles de points saillants distincts) dans un contexte similaire représenté par la forme. C'est la raison pour laquelle, ASC_{PS-PC} est plus robuste à la similarité inter-classe.

Classe #18				
Métriques	$TSLA$	ASC_{PS-PC}	$TSLA + ASC_{PS-PC}$	Apport moyen de la fusion (%)
Précision (knn=20)	64.36	72.55	75.36	6.91
MAP	0.473	0.502	0.571	8.35
Classe #19				
Métriques	$TSLA$	ASC_{PS-PC}	$TSLA + ASC_{PS-PC}$	Apport moyen de la fusion (%)
Précision (knn=20)	70.18	74.18	81.64	9.45
MAP	0.561	0.406	0.606	12.23
Classe #27				
Métriques	$TSLA$	ASC_{PS-PC}	$TSLA + ASC_{PS-PC}$	Apport moyen de la fusion (%)
Précision (knn=20)	53.00	49.82	59.45	8.05
MAP	0.317	0.271	0.341	4.66
Classe #29				
Métriques	$TSLA$	ASC_{PS-PC}	$TSLA + ASC_{PS-PC}$	Apport moyen de la fusion (%)
Précision (knn=20)	51.39	57.50	62.78	8.33
MAP	0.379	0.332	0.438	8.25

TABLE 5.2 – Apport de la fusion $TSLA + ASC_{PS-PC}$ sur les classes #18, #19, #27 et #29. La méthode de fusion IRP est utilisée.

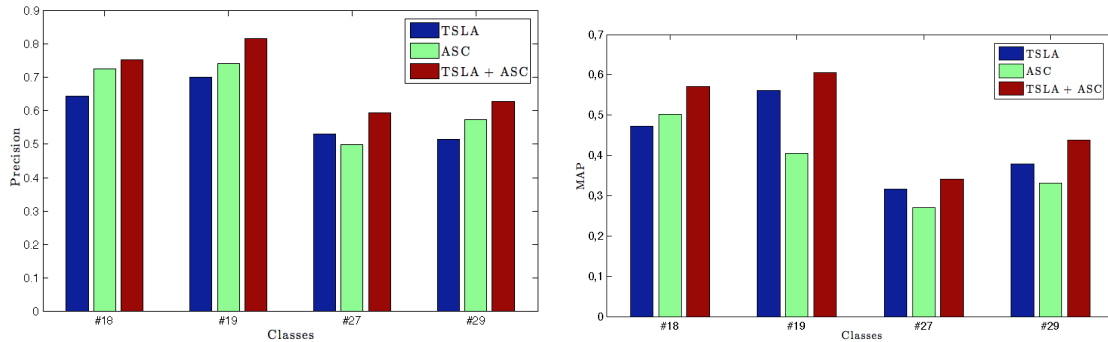


FIGURE 5.2 – Fusion des descriptions $TSLA$ et ASC_{PS-PC} sur les classes #18, #19 et #27, #29

La figure 5.3 illustre par un exemple l’apport de la fusion d’un descripteur de forme (TSL) et ASC_{PS-PC} sur une requête visuelle. Sur les cinq premières images retournées, TSL et ASC_{PS-PC} obtiennent respectivement des taux de 3/5 et 4/5 d’identification correcte. Les deux images non pertinentes retournées par le descripteur TSL sont dues au cas classique de similarité inter-classes. Lorsque le descripteur TSL est fusionné avec ASC_{PS-PC} , la description devient plus robuste et les cinq premières images appartiennent à la bonne espèce végétale.

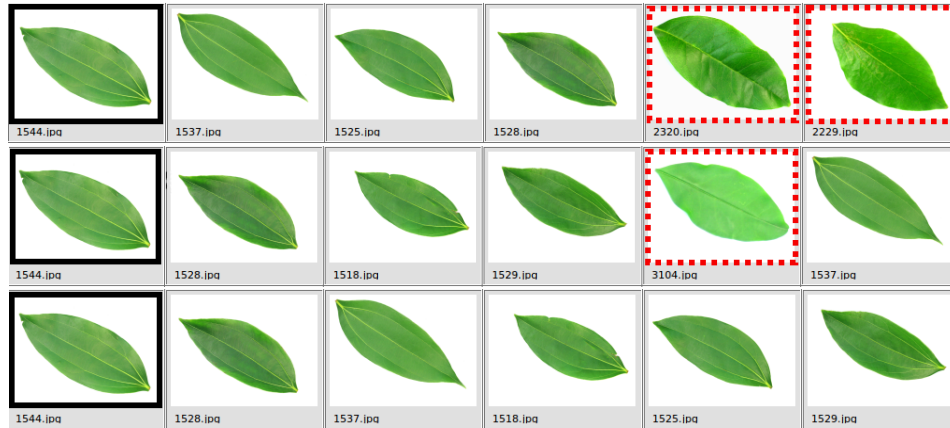


FIGURE 5.3 – Requêtes visuelles avec les descripteurs TSL (haut), ASC_{PS-PC} (milieu) et $TSL + ASC_{PS-PC}$ (bas). Les images non pertinentes retournées sont encadrées avec un trait rouge discontinu

5.2.2 Fusion des descripteurs de nervation

Descripteur	$Hough$	ASC_{PN-PN}	ASC_{PN-PC}	$Hough + ASC_{PN-PN}$	$Hough + ASC_{PN-PC}$
Précision(knn=1)	79.18%	71.73%	86.83%	93.60%	96.46%
MAP	0.558	0.378	0.538	0.547	0.664

TABLE 5.3 – Résultats individuels des descripteurs de nervation

Méthode de fusion	IRP		BC	
Descripteurs	$Hough + ASC_{PN-PN}$	$Hough + ASC_{PN-PC}$	$Hough + ASC_{PN-PN}$	$Hough + ASC_{PN-PC}$
Précision (knn=1)	83.32%	93.02%	93.60%	96.46%
MAP	0.522	0.633	0.546	0.664

TABLE 5.4 – Résultats de fusion des descripteurs de nervation

k-NN	PFT+ [KNSS11b]	$Hough$	ASC_{PN-PN}	ASC_{PN-PC}	$Hough + ASC_{PN-PN}$	$Hough + ASC_{PN-PC}$
1-NN	93.13%	79.18%	71.73%	86.83%	93.60%	96.46%
3-NN	97.81%	89.14%	83.74%	93.96%	97.53%	98.63%
5-NN	98.12%	93.60%	87.46%	96.48%	98.16%	99.03%

TABLE 5.5 – Comparaison des taux d'identification des descripteurs de nervation sur la base Flavia

Le tableau 5.5 présente les résultats obtenus par notre descripteur de nervures et ceux de l'approche PFT+ [KNSS11b] qui combine un descripteur de nervures, un descripteur de forme (Fourier polaire) et un descripteur de couleur (moments de couleur).

Nous avons suivi le même protocole d'évaluation que [KNSS11b] qui consiste à calculer le taux d'identification (ROC) sur les k premières images retournées. Comme Abdulkadir et al. ne fournissent pas les résultats individuels de chaque descripteur sur la base Flavia, la comparaison est effectuée sur le résultat donné par leur approche combinant la forme, la

nervation et la couleur.

Le descripteur de nervures $Hough + ASC_{PN-PC}$ (que nous dénommons dans la suite $HASC$) obtient le meilleur taux d'identification (supérieur à 99% sur les 5 premières images retournées). Il dépasse le résultat obtenu par l'approche PFT+ qui utilise les nervures parmi d'autres caractères. Nous pouvons noter : - L'efficacité de la fusion de la description des directions des nervures via $Hough$ et de la description spatiale. L'apport de la fusion est net :

- $Hough + ASC_{PN-PN}$: $\approx 21\%$ par rapport à ASC_{PN-PN} et $\approx 14\%$ par rapport à $Hough$ pour la première image retournée.
- $Hough + ASC_{PN-PC}$ ($HASC$) : $\approx 10\%$ par rapport à ASC_{PN-PC} et $\approx 17\%$ par rapport à $Hough$ pour la première image retournée.
- La robustesse du descripteur $HASC$ face au bruit et aux discontinuités du réseau de nervures extrait. Prenons le cas concret de la classe #22. Les feuilles de cette espèce se caractérisent par une nervation très faiblement contrastée et la méthode d'extraction de nervures produit souvent un réseau bruité et non-connecté sur ces feuilles (cf. Figure 5.4). Par conséquent, c'est une espèce végétale difficilement identifiable via la nervation de ses feuilles. La méthode PFT+ [KNSS11b] obtient 10% d'identification correcte sur cette classe à 1-NN alors que $HASC$ atteint le score de 98.18%. Cet excellent résultat s'explique par le fait que la méthode de description de nervation $HASC$ a une très bonne insensibilité au bruit due en partie au système de vote de points d'une droite dans la description de $Hough$. Le descripteur $HASC$ permet la description du réseau de nervures bruité de la feuille présentée dans la figure 5.4 avec une précision élevée.

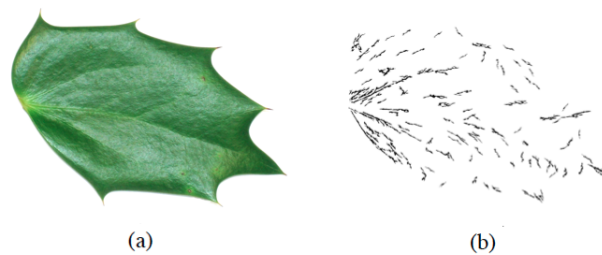


FIGURE 5.4 – Réseau de nervures bruité et non connecté. (a) Image originale ; (b) nervures extraites

Comparons maintenant les performances des descriptions $Hough$, ASC_{PN-PC} et $HASC$ par classe sur la base Flavia. Nous avons interprété les cas où ASC_{PS-PC} est plus performant que $Hough$ et vice versa.

Nous pouvons conclure que le descripteur $HASC$ est adapté à la description de nervures.

Il a montré une bonne robustesse au bruit et aux discontinuités des réseaux extraits. Ainsi, certains défauts d'extraction de nervures seront tolérés et moins contraignants lors de la phase de description. Nous avons prouvé que l'utilisation de la nervation dans le processus d'identification est envisageable et peut conduire à des résultats comparables à l'état de l'art sur la base Flavia. Le descripteur *HASC*, fusion d'une description spatiale et d'une description des principales directions des nervures a montré son efficacité sur l'ensemble de la base Flavia.

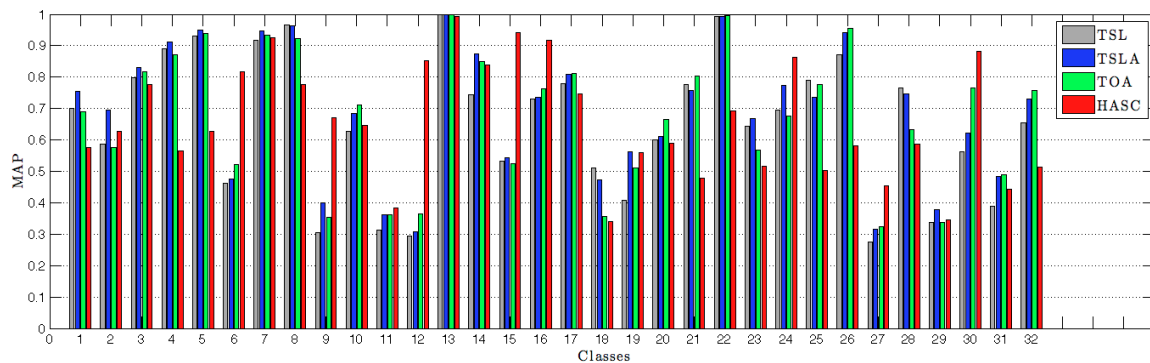
En revanche, la procédure d'extraction et de description de la nervation dépend étroitement de la qualité des images traitées. En fait, un contraste non suffisant des nervures peut conduire à une forte baisse de performance même dans le cas des images de bonne qualité comme celles de la base Flavia. Malgré un bon score d'identification, la nervation reste moins discriminante que la forme de la feuille. De plus, le contour est beaucoup plus facile à extraire que le réseau de nervures dans le cas des scans et des pseudo-scans de feuilles. Néanmoins, l'utilisation de l'information fournie par la nervation est utile pour compléter la description de forme. En effet, les feuilles de différentes espèces qui ont des formes similaires sont plus facilement différenciables grâce à leurs nervures (classes #1, #8, #9, #10).

5.3 FUSION DE LA DESCRIPTION DE LA FORME ET DE LA NERVATION

Dans cette section, la feuille est décrite par sa forme et ses nervures. Cela nous permet d'étudier plusieurs aspects :

- la pertinence de l'utilisation conjointe de ces deux caractères pour l'identification des espèces
- l'apport de la nervation dans des cas de similarité de forme inter-espèce
- l'évaluation de notre approche d'extraction et de description de nervures en terme de capacité à apporter une meilleure distinction entre les différentes espèces
- la comparaison avec les méthodes de la littérature qui combinent la forme et la nervation

Nous nous proposons tout d'abord de comparer les résultats obtenus par les descripteurs triangulaires de forme et le descripteur de nervures.

FIGURE 5.5 – Comparaison de *HASC* et les descripteurs triangulaires de forme, par classe sur la base Flavia

5.3.1 Comparaison entre les descriptions de forme et de nervures

Rappelons les résultats du descripteur de nervures *HASC* et ceux des descripteurs de formes proposés.

Descripteurs	Précision ($knn = 20$)	MAP
<i>TSLA</i>	80.11	0.659
<i>TSLA</i>	82.91	0.697
<i>TOA</i>	82.35	0.683
<i>HASC</i>	84.44	0.664

TABLE 5.6 – Résultats des descripteurs triangulaires et du descripteur de nervation *HASC* sur Flavia

Dans une approche d'identification des plantes, il est commun d'attribuer à la forme de la feuille beaucoup plus d'importance qu'à sa nervation. Le nombre de travaux de la littérature qui exploitent chacun de ces caractères confirme cette observation. La difficulté principale qui fait que la nervation est moins utilisée, est son extraction d'une façon précise. Par conséquent, les méthodes de description de nervures ne sont pas très variées comme c'est le cas des descripteurs de forme. Cela a été la motivation essentielle pour consacrer une partie de nos travaux pour développer une approche d'identification des espèces végétale utilisant l'information fournie par la nervation. A notre connaissance, il n'existe pas de travaux qui fournissent des résultats d'un descripteur de nervures sur une base publique. Les résultats du tableau 5.6 confirment que la nervation est capable d'obtenir des résultats d'identification comparables avec les descripteurs de forme, du moins, sur la base Flavia. Ceci peut s'expliquer par le fait que *HASC* tient compte du contexte du contour dans la description spatiale des points de nervation.

Le descripteur *HASC* obtient une meilleure précision devant tous les descripteurs triangulaires de forme. En terme de MAP, il est positionné derrière *TSLA* et *TOA*.

La figure 5.5 montre les résultats obtenus par classe. *HASC* dépasse tous les triangulaires sur 9 classes sur 32. Pour la majorité de ces classes, nous remarquons un écart important entre *HASC* et le reste des descripteurs. Nous examinons les images de ces classes et les réponses des requêtes qu'elle donnent. Nous en déduisons qu'il s'agit encore une fois d'un cas de la similarité inter-classe (cf. Figure 5.6).

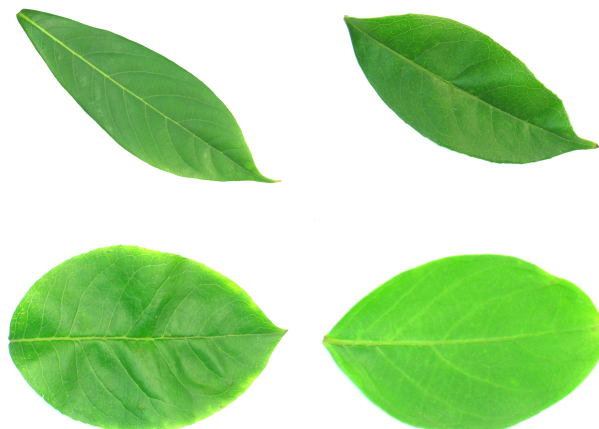


FIGURE 5.6 – Formes similaires des feuilles des classe #6 et #12 et des classes #9 et #15 de la base Flavia

La figure 5.5 montre bien que les résultats de *HASC* et des descripteurs triangulaires sont dé-corrélés vu que l'information décrite est différente. Dans le paragraphe suivant, nous allons vérifier la complémentarité des descripteurs de forme et de nervation lorsqu'ils sont fusionnés.

5.3.2 Fusion des descripteurs de forme et de nervation

Le descripteur de nervures *HASC* a été évalué sur la base Flavia. Donc, la fusion avec les descripteurs de forme est étudiée uniquement sur cette base. Les résultats de fusion sont reportés dans le tableau 5.7.

Descripteurs	<i>IRP</i>	<i>BC</i>
<i>TSL + HASC</i>	0.716	0.756
<i>TSLA + HASC</i>	0.726	0.760
<i>TOA + HASC</i>	0.718	0.755

TABLE 5.7 – MAP de la fusion des descripteurs triangulaires de forme avec le descripteur de nervures *HASC*

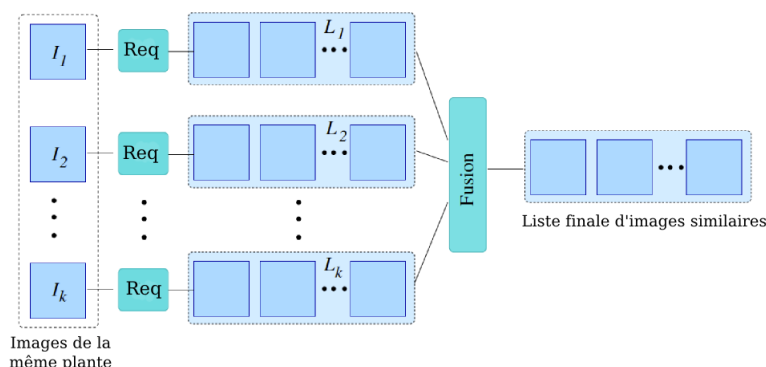


FIGURE 5.7 – Processus de la méthode des requêtes multiples

5.4 REQUÊTES MULTIPLES

Dans cette section nous introduisons une nouvelle méthode d'évaluation des descripteurs via des requêtes multiples. Le principe est le suivant : au lieu de considérer le résultat d'une image singulière pour trouver l'espèce, nous considérons le résultat donné par la soumission d'un ensemble d'images provenant d'un même individu (i.e un même arbre) comme requête). Les listes des images similaires sont calculées pour chaque image requête. Elles sont ensuite fusionnées en une liste finale. (cf. Figure 5.7). La méthode des requêtes multiples suppose qu'une information sur les individus auxquels appartiennent les images est fournie. C'est le cas des bases ImageCLEF [GBJ⁺11, GBJ⁺12, GJB⁺13].

Les images des feuilles d'un même arbre peuvent avoir des caractéristiques visuellement différentes à cause de la variabilité intra-espèce. Nous espérons, avec l'application des requêtes multiples, améliorer l'identification des espèces végétales de façon à ce que chaque image d'une plante contribue à l'identification.

Dans la suite, nous étudions l'apport de la méthode des requêtes multiples sur la base ImageCLEF 2012.

5.4.1 Évaluation sur ImageCLEF 2012

Résultats des descripteurs individuels

Le score d'identification S est calculé en utilisant des requêtes multiples sur les scans et les pseudo-scans. Nous le comparons au score obtenu avec des requêtes simples dans les tableaux 5.8 et 5.9. Pour tous les descripteurs, les requêtes multiples permettent d'améliorer le score d'identification.

Prenons le cas des scans. L'amélioration de la performance de TSL , $TSLA$ et TOA est du



FIGURE 5.8 – Exemples des pseudo-scans de test dans la base ImageCLEF.

même ordre de grandeur pour les trois descripteurs. Dans le cas de ASC_{PS-PC} , l'apport est plus important. De plus, ASC_{PS-PC} obtient la moins bonne performance avec des requêtes simples mais devient le meilleur dans le cas des requêtes multiples. L'information fournie par chaque image d'un arbre donné, sur la corrélation spatiale entre ses points saillants et son contour semble être plus utile que celle sur son contour seul, dans le cas des scans. Sur les pseudo-scans, l'amélioration apportée par les requêtes multiples est assez importante pour les descripteurs triangulaires alors qu'elle est relativement faible pour ASC_{PS-PC} . Cela confirme deux observations :

- La pertinence de la description de la forme dans le cas des pseudo-scans par leur forme. Les résultats avec les requêtes simples le montrent bien et les résultats avec des requêtes multiples le confirment.
- les points saillants ne constituent pas un caractère déterminant d'identification des pseudo-scans de feuilles. En effet, certaines images de la base de test contiennent des feuilles mortes où il n'existe pas d'éléments saillants (nervation, texture, etc.) ou des feuilles avec une luminance irrégulière.

L'évaluation de la méthode des requêtes multiples a montré qu'elle est capable d'améliorer les scores d'identification des descripteurs individuels. Le meilleur score d'identification des scans et des pseudo-scans est obtenu respectivement par ASC_{PS-PC} et $TSLA$.

Examinons maintenant si la fusion des descripteurs suivant la méthode de requêtes multiples va améliorer les processus d'identification. Ceci est discuté dans le paragraphe suivant.

Descripteur	Requêtes simples	Requêtes multiples (RM)	Apport des requêtes multiples
<i>TSL</i>	0.521	0.587	6,6%
<i>TSLA</i>	0.518	0.581	6,3%
<i>TOA</i>	0.537	0.598	6,1%
<i>ASC_{PS-PC}</i>	0.510	0.615	10.5%

TABLE 5.8 – Résultats des descripteurs en utilisant les requêtes multiples sur les scans d'ImageCLEF 2012

Descripteur	Requêtes simples	Requêtes multiples (RM)	Apport des requêtes multiples
<i>TSL</i>	0.556	0.690	13.4%
<i>TSLA</i>	0.612	0.789	17.7%
<i>TOA</i>	0.622	0.718	9.6%
<i>ASC_{PS-PC}</i>	0.409	0.454	4.45%

TABLE 5.9 – Résultats des descripteurs en utilisant les requêtes multiples sur les pseudo-scans d'ImageCLEF 2012

Fusion des descriptions foliaires

Stratégies de fusion

Soient Q_1, \dots, Q_q q images de feuilles provenant du même arbre et constituant une requête multiple et d_1, \dots, d_k k descripteurs à combiner.

Il existe deux stratégies de fusion si plusieurs descripteurs sont combinés (cf. Figure 5.9) :

- A. Pour chaque image requête Q_i , k listes d'images similaires sont calculées avec k descripteurs. Ces k listes sont combinées avec une méthode de fusion choisie parmi celles décrites dans la section 5.1. Les q listes d'images résultantes elles sont alors fusionnées.
- B. Pour chaque descripteur d_i pris parmi la liste des descripteurs d_1, \dots, d_k , une liste d'images similaires L_i est établie à partir de la fusion des q listes de similarité calculées avec d_i sur les images Q_1, \dots, Q_q . Une liste d'images correspondant à la requête multiple Q_1, \dots, Q_q sur la base en utilisant le descripteur d_i est donc construite. Ces listes sont alors combinées (cf. Figure 5.9.B)

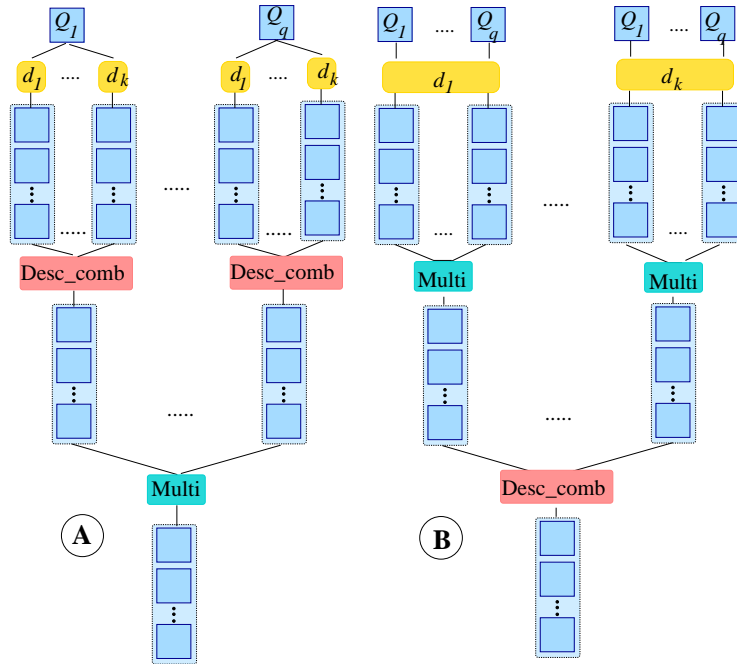


FIGURE 5.9 – Les deux stratégies de fusion des descripteurs dans le cas de requêtes multiples. 'Multi' désigne la fusion des listes des requêtes multiples. 'Desc_comb' signifie la combinaison des descripteurs

Résultats de fusion

Descripteur	Fusion A					
	IRP	IRPCombMNZ	BC	BCCombMNZ	MP	WPr
$TSL + ASC_{PS-PC}$	0.640	0.653	0.628	0.615	0.476	0.568
$TSLA + ASC_{PS-PC}$	0.636	0.671	0.657	0.653	0.486	0.551
$TOA + ASC_{PS-PC}$	0.657	0.655	0.663	0.647	0.476	0.593

TABLE 5.10 – Résultats de fusion avec des requêtes multiples sur les scans d'ImageCLEF 2012. La méthode de fusion A est utilisée

Descripteur	Fusion B					
	IRP	IRPCombMNZ	BC	BCCombMNZ	MP	WP
$TSL + ASC_{PS-PC}$	0.624	0.649	0.624	0.672	0.626	0.651
$TSLA + ASC_{PS-PC}$	0.633	0.621	0.621	0.654	0.559	0.614
$TOA + ASC_{PS-PC}$	0.640	0.644	0.646	0.609	0.599	0.623

TABLE 5.11 – Résultats de fusion avec des requêtes multiples sur les scans d'ImageCLEF 2012. La méthode de fusion B est utilisée

Descripteur	Fusion A					
	<i>IRP</i>	<i>IRPCombMNZ</i>	<i>BC</i>	<i>BCCombMNZ</i>	<i>MP</i>	<i>WP</i>
$TSL + ASC_{PS-PC}$ (RM)	0.723	0.676	0.674	0.615	0.561	0.632
$TSLA + ASC_{PS-PC}$ (RM)	0.742	0.735	0.732	0.730	0.602	0.639
$TOA + ASC_{PS-PC}$ (RM)	0.740	0.737	0.736	0.682	0.628	0.699

TABLE 5.12 – Scores d’identification de la fusion des descripteurs en utilisant des requêtes multiples sur les pseudo-scans d’ImageCLEF 2012. La méthode de fusion A est utilisée

Descripteur	Fusion B					
	<i>IRP</i>	<i>IRPCombMNZ</i>	<i>BC</i>	<i>BCCombMNZ</i>	<i>MP</i>	<i>WP</i>
$TSL + ASC_{PS-PC}$ (RM)	0.724	0.714	0.592	0.596	0.627	0.681
$TSLA + ASC_{PS-PC}$ (RM)	0.791	0.730	0.674	0.660	0.707	0.624
$TOA + ASC_{PS-PC}$ (RM)	0.739	0.740	0.676	0.682	0.659	0.619

TABLE 5.13 – Scores d’identification de la fusion des descripteurs en utilisant des requêtes multiples sur les pseudo-scans d’ImageCLEF 2012. La méthode de fusion B est utilisée

5.4.2 Évaluation sur ImageCLEF 2013

Pour fixer les paramètres des descripteurs triangulaires, le même nombre de points échantillonnés sur le contour $N = 400$ est utilisé et nous avons fait varier le reste des paramètres : le nombre d’échelles N_s , la longueur de la portion du contour décrite n et la fonction de distance entre deux triangles successifs $d^{lin/log}$. Les scores d’identification des trois descripteurs sont calculés en faisant varier ces paramètres. Les résultats sont donnés par le tableau 5.14.

Descripteur	$N_s = 10, n = 50, d(k) = 5 \times k$	$N_s = 10, n = 100, d(k) = 10 \times k$	$N_s = 20, n = 40, d(k) = 2 \times k$
TSL	0.548	0.539	0.557
$TSLA$	0.569	0.541	0.626
TOA	0.582	0.535	0.538

TABLE 5.14 – Scores d’identification des descripteurs triangulaires avec une variation de paramètres

Pour l’évaluation du descripteur ASC_{PS-PC} , nous avons fait varier la taille de l’ensemble de calcul. Pour des paramètres de 50 puis 100, ASC_{PS-PC} obtient respectivement un score de 0.364 et de 0.478.

Dans la suite, les meilleurs paramètres pour chaque descripteur sont retenus. L’étape suivante est consacrée la comparaison des résultats données par la méthode des requêtes multiples, reportés dans le tableau 5.15.

Comme sur la base ImageCLEF 2012, l’apport de la méthode des requêtes multiples est positif pour l’ensemble des descripteurs. L’apport le plus important est dans le cas du descripteur ASC_{PS-PC} . $TSLA$ obtient le meilleur score d’identification (0.65) en utilisant les requêtes multiples sur cette base. Il est à noter que le score de $TSLA$ (avec des requêtes

simples ou multiples) dépasse toutes les méthodes de l'état de l'art évaluées sur la base ImageCLEF 2013 [GJB⁺13] où le meilleur score de la littérature est de 0.607.

Descripteur	Requêtes simples	Requêtes multiples (RM)	Apport des requêtes multiples
<i>TSL</i>	0.557	0.583	2.6%
<i>TSLA</i>	0.626	0.65	2.4%
<i>TOA</i>	0.582	0.607	2.5%
<i>ASC_{PS-PC}</i>	0.478	0.571	9.3%

TABLE 5.15 – Scores d'identification en utilisant les requêtes multiples sur la base ImageCLEF 2013

Le tableau 5.16 montre les résultats de fusion de *ASC_{PS-PC}* et les descripteurs triangulaires de forme. Les méthodes de fusion *BCCombMNZ* et *BC* sont les meilleures en tenant compte des trois configurations. Le meilleur score d'identification est obtenu par la fusion de *TSLA* et *ASC_{PS-PC}* avec la méthode *BC*. Cette fusion améliore les scores individuels obtenus avec des requêtes simples mais n'améliore pas le score de *TSLA* obtenu avec les requêtes multiples.

Pour la fusion des descripteurs en utilisant les requêtes multiples, nous avons utilisé la stratégie de fusion A car elle est légèrement meilleure sur ImageCLEF 2012. De plus, les deux stratégies ont des performances comparables. Les résultats reportés dans le tableau 5.17 confirment l'utilité de l'approche d'identification avec les requêtes multiples. Tous les scores se sont améliorés par rapport à la fusion avec des requêtes simple. Nous arrivons à dépasser largement la meilleure performance de la littérature grâce à la fusion *TSLA* + *ASC_{PS-PC}*. (0.683 contre 0.607).

Les différentes expérimentations effectuées permettent de constater aussi que les méthodes de fusion *IRPCombMNZ*, *BC* et *BCCombMNZ* se comportent généralement mieux que les autres méthodes.

Descripteur	<i>IRP</i>	<i>BCCombMNZ</i>	<i>BC</i>	<i>BCCombMNZ</i>	<i>MP</i>
<i>TSL</i> + <i>ASC_{PS-PC}</i>	0.583	0.593	0.596	0.601	0.58
<i>TSLA</i> + <i>ASC_{PS-PC}</i>	0.625	0.638	0.646	0.644	0.623
<i>TOA</i> + <i>ASC_{PS-PC}</i>	0.598	0.602	0.611	0.623	0.591

TABLE 5.16 – Scores d'identification de la fusion des descripteurs sur la base ImageCLEF 2013

Descripteur	IRP	IRPCombMNZ	BC	BCCombMNZ	MP
$TSL + ASC_{PS-PC}$ (RM)	0.608	0.613	0.611	0.618	0.626
$TSLA + ASC_{PS-PC}$ (RM)	0.671	0.683	0.683	0.669	0.664
$TOA + ASC_{PS-PC}$ (RM)	0.618	0.633	0.653	0.659	0.621

TABLE 5.17 – Scores d’identification de la fusion des descripteurs en utilisant les requêtes multiples sur la base ImageCLEF 2013

5.5 COMPARAISON AVEC L’ÉTAT DE L’ART

Dans cette section, nous comparons nos meilleurs résultats d’identification avec ceux de l’état de l’art sur trois bases publiques. Nous gardons pour chaque base, les meilleures approches et les meilleurs paramètres discutés dans les sections précédentes.

5.5.1 Comparaison sur la base ImageCLEF 2012

Sur la base ImageCLEF 2012, la fusion des descripteurs triangulaires de forme et du descripteur ASC_{PS-PC} s’avère extrêmement efficace. Le tableau 5.18 montre que la fusion associée à des requêtes multiples (RM) dépasse largement les scores des méthodes évaluées sur les deux catégories d’images scans et pseudo-scans d’ImageCLEF2012.

	Scans	Pseudo-scans
3 meilleurs scores des participants	0.58	0.59
	0.49	0.55
	0.47	0.54
$TSL + ASC_{PS-PC}$ (RM)	0.672	0.724
$TSLA + ASC_{PS-PC}$ (RM)	0.671	0.791
$TOA + ASC_{PS-PC}$ (RM)	0.663	0.740

TABLE 5.18 – Comparaison des scores d’identification sur la base ImageCLEF 2012

5.5.2 Comparaison sur la base ImageCLEF 2013

Considérons la tâche d’identification de plante d’ImageCLEF 2013. Notons que nous avons participé à cette tâche avec 4 méthodes : Inria PlantNet Run 1, 2, 3 et 4. Toutes les méthodes ont été positionnées derrière la meilleure performance obtenue dans *Sabancı Okan Run 1* (se référer à [GJB⁺13] pour les détails).

Avec un choix approprié des paramètres (cf. 4.6, page 144), il s’est avéré que plusieurs de nos approches arrivent à dépasser le meilleur score obtenu dans ImageCLEF 2013. Les méthodes sont listées dans le tableau 5.19.

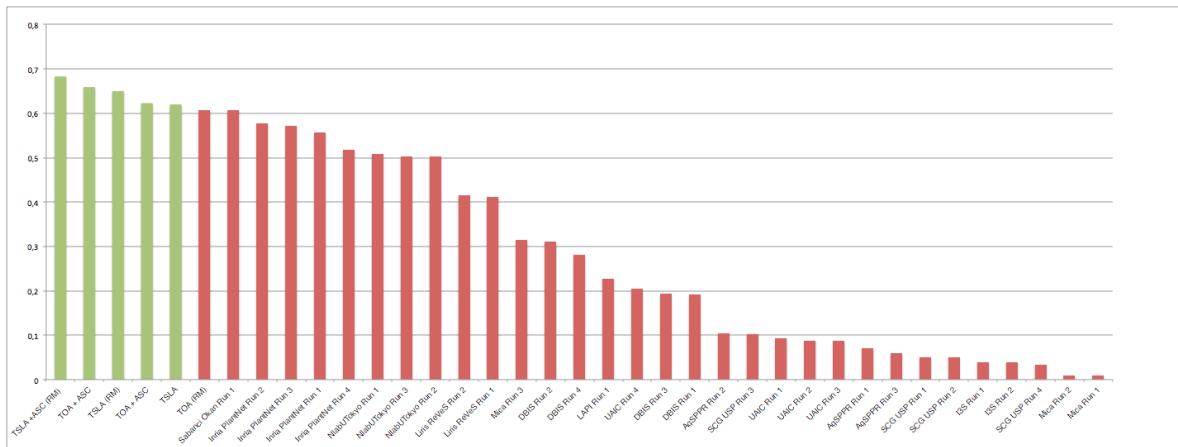


FIGURE 5.10 – Comparaison avec les méthodes ayant participé ImageCLEF2013

Le meilleur score d'identification est obtenu par la fusion $TSLA + ASC_{PS-PC}$ en suivant la méthodes des requêtes multiples (RM). La figure 5.10 montre un aperçu sur le positionnement des nos meilleures approches sur ImageCLEF 2013, par rapport aux autres approches.

Nos résultats	
$TSLA + ASC_{PS-PC}$ (RM)	0.683
$TOA + ASC_{PS-PC}$	0.659
$TSLA$ (RM)	0.65
$TOA + ASC_{PS-PC}$	0.623
$TSLA$	0.62
TOA (RM)	0.607
Dix meilleurs scores des participants à ImageCLEF 2013	
Sabancı Okan Run 1	0.607
Inria PlantNet Run 2	0.577
Inria PlantNet Run 3	0.572
Inria PlantNet Run 1	0.557
Inria PlantNet Run 4	0.517
NlabUTokyo Run 1	0.509
NlabUTokyo Run 3	0.502
NlabUTokyo Run 2	0.502
Liris ReVeS Run 2	0.416
Liris ReVeS Run 1	0.412

TABLE 5.19 – Comparaisons avec les résultats de l'état de l'art sur la base ImageCLEF2013. (RM) correspond aux requêtes multiples

5.5.3 Comparaison sur la base Flavia

Nous évaluons dans cette expérimentation la performance de la fusion du descripteur de forme $TSLA$ et le descripteur de nervure $HASC$. La comparaison est effectuée avec la

méthode PFT dans [KNSS11b] qui combine aussi un Fourier polaire sur la forme avec un descripteur de nervures se basant sur le calcul du rapport entre la taille des nervures en terme du nombre de pixels et celle du contour. Nous avons suivi la métrique d'évaluation de [KNSS11b] qui consiste à calculer le taux d'identification (ROC) sur de multiples $k - NN$. Les résultats de comparaison figurent dans le tableau 5.20. Nous arrivons à dépasser les résultats de [KNSS11b] en utilisant le descripteur de nervure *HASC* individuellement et en effectuant la fusion *TSLA + HASC*. Le taux d'identification est très proche de 100% si on considère les cinq premières images retournées. Cela montre de nouveau l'efficacité de notre démarche globale.

k-NN	PFT+ [KNSS11b]	<i>HASC</i>	<i>TSLA + HASC</i>
1-NN	93.13%	96.46%	97.27%
3-NN	97.81%	98.63%	99.84%
5-NN	98.12%	99.03%	99.89%

TABLE 5.20 – Comparaison des taux d'identification de *TSLA+HASC* avec la méthode dans [KNSS11b] qui combine la description de forme et de nervation sur la base Flavia

Une récente étude [KSL14] propose d'expérimenter un ensemble de descripteurs de forme sur la base Flavia. Ces descripteurs sont testés individuellement puis en effectuent des fusions. Nous désignons par AP dans la suite l'approche proposée dans [KSL14]. Les auteurs calculent la valeur de MAP comme métrique d'évaluation. Nous comparons les résultats individuels et fusionnés obtenus avec les nôtres respectivement dans les tableaux 5.22 et 5.21.

Nous remarquons que dans le cas des descripteurs individuels, tous nos descripteurs dépassent les résultats de l'état de l'art. ASC_{PS-PC} , *HASC* et les descripteurs triangulaires (*TOA* et *TSLA*), bien qu'il décrivent des caractéristiques différentes de la feuille, obtiennent des valeurs de MAP meilleures que celles dans [KSL14]. L'utilisation implicite de la forme dans ASC_{PS-PC} et *HASC* à travers un vote du contour a bien montré son utilité quand ces descripteurs sont comparés à l'état de l'art. Notre meilleur descripteur de forme est *TSLA* ; il obtient une MAP largement supérieure à la meilleure MAP dans [KSL14].

Pour la comparaison des résultats de la fusion, nous tirons les deux remarques suivantes :

- L'apport de la fusion est plus important sur nos descripteurs : 6.29% contre 3.34% si nous comparons la différence entre les meilleures valeurs de MAP (individuelle et de la fusion).
- Les résultats de fusion montrent un grand écart avec ceux de [KSL14]. Ceci peut s'expliquer par le fait que les auteurs combinent essentiellement des descripteurs de

contour alors que dans notre cas les descripteurs fusionnés décrivent différents caractères de la feuille (forme, points saillants et nervation). Cette fusion de descriptions des différents caractères permet de résoudre certains cas d'inter-similarité de forme comme nous l'avons vu précédemment.

Kurtek et. al (2014) [KSL14]						
Méthodes	D2	AF	MSDM	GEDT	AP	SM
MAP	0.421	0.458	0.47	0.4708	0.523	0.554
Nos résultats						
Nos descripteurs	ASC_{PS-PC}	TSL	$HASC$	TOA	$TSLA$	
MAP (%)	0.587	0.659	0.664	0.683	0.697	

TABLE 5.21 – MAP de nos descripteurs individuels et MAP de ceux l'état de l'art sur la base Flavia. D2, AF, MSDM, GEDT, AP et SM sont décrits dans [KSL14]

Kurtek et. al (2014) [KSL14]					
Méthodes	AP+D2	AP+MSDM	AP+GEDT	AP+AF	AP+SM
MAP	0.537	0.538	0.546	0.558	0.587
Nos résultats					
Nos résultats	$TSLA + ASC_{PS-PC}$	$TOA + HASC$	$TSL + HASC$	$TSLA + HASC$	
MAP	0.728	0.755	0.756	0.760	

TABLE 5.22 – MAP de la fusion de nos descripteurs et MAP de la fusion des descripteurs de l'art sur la base Flavia

5.6 CONCLUSION

En conclusion, notre approche d'identification s'avère simple et efficace. La fusion des différentes descriptions foliaires améliorent généralement l'identification. Ceci a été confirmé expérimentalement. Nous avons de plus étudié l'apport de chaque descripteur dans l'identification et avons identifié ses limites.

Après tout ce travail, nous pouvons constater que la tâche d'identification des espèces végétales est complexe. Le fait de choisir de décrire un unique caractère montre rapidement ses limites. Finalement, ces expérimentations permettent de donner quelques recommandations :

La forme de la feuille reste le caractère de base d'un système d'identification de plantes. Dans le cadre de nos travaux, nous recommandons l'utilisation du descripteur $TSLA$ car il a montré une bonne stabilité de performance sur les différentes bases de feuilles. Tous les descripteurs de forme sont mis en échec par la similarité inter-espèce. Le contexte botanique est très particulier car ce cas peut représenter une réelle difficulté même pour les experts.

Il est fréquent que la similarité inter-classe ne se résume pas à une simple ressemblance de formes mais représente des formes parfaitement identiques. Dans ce cas précis, nous pensons qu'il est inutile d'augmenter la robustesse du descripteur mais tout simplement, il faut penser à enrichir la description avec un autre caractère foliaire afin de lever cette ambiguïté.

Chapitre 6

Conclusion générale et perspectives

Cette thèse a été consacrée à l'identification automatique des espèces végétales à partir de la donnée d'une ou de plusieurs images de feuilles issues d'un même arbre. Elle s'est effectuée dans le cadre du projet Pl@ntNet et les développements ont été intégrés dans le moteur de recherche par le contenu IKONA-Maestro [BFF⁺01].

Plus précisément, nous avons proposé de nouvelles descriptions foliaires robustes en suivant une méthode de description géométrique locale pour la majorité des descripteurs :

- Un ensemble de représentations triangulaires multi-échelle du contour de la feuille a été proposé. Trois types de descriptions des triangles ont été étudiés et comparés (TSL, TSLA, TOA). Ces descriptions sont robustes à l'occultation partielle et aux déformations locales du contour.
- Un deuxième groupe de descripteurs permet de décrire les corrélations entre les points du contour (ASC_{PC-PC}), entre les points saillants (ASC_{PS-PS}) ou entre les points saillants et le contour (ASC_{PS-PC}). Cette dernière description a obtenu les meilleurs scores par rapport à toutes les descriptions de ce groupe sur l'ensemble des bases évaluées. Néanmoins, il est à noter que cette description est sensible au bruit donc à la qualité de l'image de départ.
- La représentation du réseau de nervures a conduit à l'introduction de deux nouveaux descripteurs dérivés du contexte de forme : ASC_{PN-PN} qui représente les relations spatiales

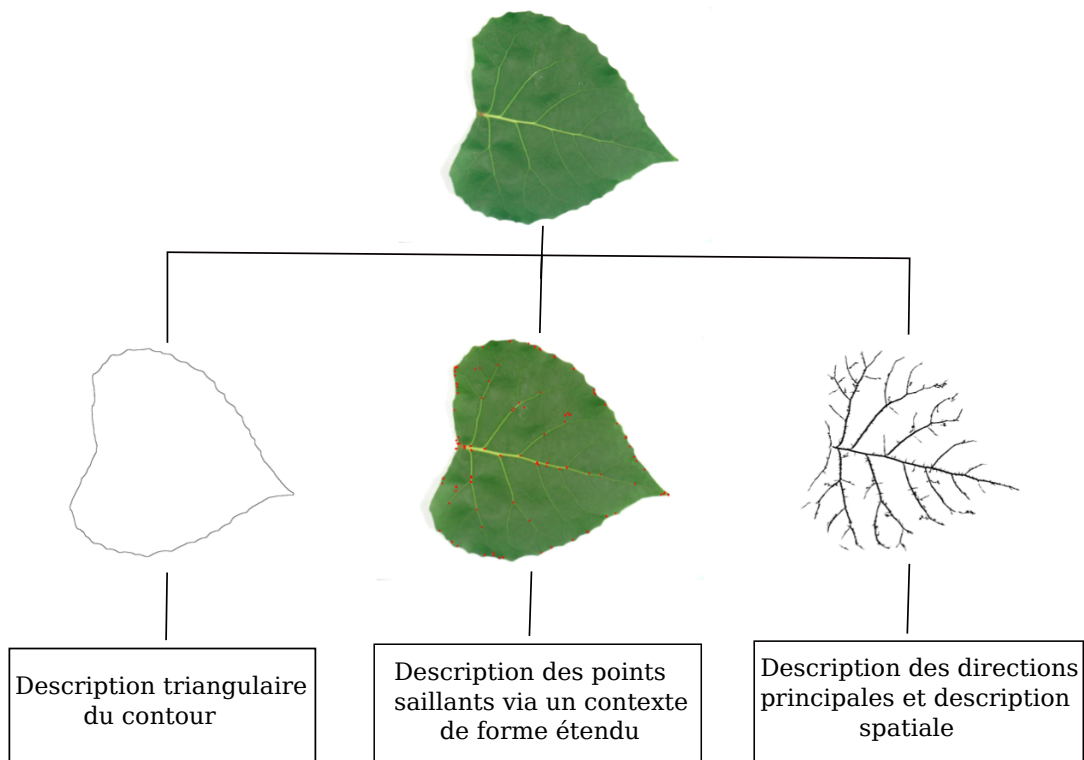


FIGURE 6.1 – Résumé visuel des descriptions proposées

des points de nervures entre eux et ASC_{PN-PC} qui représente la corrélation entre les points de nervures et le contour de la feuille. Un descripteur global du réseau de nervures a été également proposé (*Hough*). Il vise à détecter les directions principales des nervures.

L'évaluation sur la base Flavia a révélé que le descripteur *Hough* pris isolément obtenait le meilleur score par rapport aux deux autres.

Une étude expérimentale a été effectuée sur six bases d'images de feuilles publiques. Le tableau 6.1 montre les rangs respectifs des descripteurs proposés sur ces bases.

Bases d'images	Swedish	Flavia	Leaf100	IC 2011	IC 2012	IC 2013
Descripteurs de forme						
<i>TSL</i>	3	4	3	3	3	2
<i>TSLA</i>	1	2	1	1	2	1
<i>TOA</i>	2	3	2	2	1	3
Contexte de forme étendu (<i>ASC</i>)						
ASC_{PC-PC}	5	6	-	4	4	4
ASC_{PS-PS}	7	7	-	7	-	-
ASC_{PS-PC}	4	5	-	5	5	5
$eASC_{PS-PC}$	6	1	-	6	6	6
Descripteur de nervation						
<i>Hough</i>	-	8	-	-	-	-
ASC_{PN-PN}	-	10	-	-	-	-
ASC_{PN-PC}	-	9	-	-	-	-

TABLE 6.1 – Rangs des descripteurs individuels développés dans cette thèse sur les six bases utilisées

Cette étude nous a permis de recommander l'utilisation de la description *TSLA* pour décrire la forme des feuilles avec les paramètres suivants :

- $N = 400$, $n = 50$, $N_s = 10$ (d_5^{lin}).
- Une deuxième configuration a obtenu des scores comparables sur les bases Swedish et Flavia :
 - $N = 400$, $n = 64$, $N_s = 6$ (d_2^{log}).

Les fusions tardives des descripteurs proposés ont permis d'améliorer les scores d'identification. Sur toutes les bases où la fusion tardive a été évaluée (Flavia, IC2012 et IC2013), l'apport de la fusion a été clair notamment avec l'utilisation des requêtes multiples (RM). Sur la majeure partie de l'ensemble des bases évaluées, nos descripteurs (individuels ou fusionnés) ont pu dépasser les meilleurs scores de la littérature (cf. Tableau 6.2).

Bases	Métrique	littérature	Nos résultats	Meilleures descriptions
Swedish	Précision 1-NN	97.92 [WR08]	96.53	<i>TSLA</i>
Flavia	Précision 1-NN	93.13 [KNSS11b]	97.27	<i>TSLA + HASC</i>
	MAP	0.587 [KSL14]	0.760	<i>TSLA + HASC</i>
IC2011	Score <i>S</i> (scans)	0.778 [YMB12]	0.802	<i>TSL</i>
	Score <i>S</i> (pseudo-scans)	0.725 [YMB12]	0.780	<i>TOA</i>
IC2012	Score <i>S</i> (scans)	0.58 [YAT12]	0.67	<i>(TSLA ou TSLA)+ASC_{PS-PC}</i> (RM)
	Score <i>S</i> (pseudo scans)	0.59 [BYM ⁺ 12]	0.79	<i>TSLA + ASC_{PS-PC}</i> (RM)
IC2013	Score <i>S</i>	0.607 [YAY13]	0.683	<i>TSLA + ASC_{PS-PC}</i> (RM)

TABLE 6.2 – Comparaison de nos meilleurs résultats avec ceux de la littérature

APPLICABILITÉ DES APPROCHES PROPOSÉES

Les objectifs mentionnés en introduction ont été globalement atteints. En particulier, nos descriptions triangulaires multi-échelle se sont avérées efficaces pour différencier des formes délimitées par un contour fermé et assez proches les unes des autres.

Le cas de formes génériques 2D serait à étudier. Ce type de descripteurs pourrait aussi être utilisé dans d'autres bases comme celle des fleurs d'orchidées de Tunisie présentée dans [RSBG13]. Le cas d'un ensemble plus général d'images de fleurs est différent. En effet, la couleur est un caractère fortement discriminant comme l'ont noté Nilsback et Zisserman [NZ08]. De plus, la fleur est composée de plusieurs éléments et elle n'est pas forcément plane. Cependant, une fusion des descripteurs introduits dans cette thèse pourrait être évaluée sur des images de fleurs.

L'approche morphologique d'extraction de nervures est générique et applicable à tout type de réseaux. Une approche similaire a permis la détection de vaisseaux sanguins rétiniens dans une image [Wal03] (cf. Figure 6.2). Elle pourrait également être transposée pour obtenir la nervation alaire d'insectes à partir d'images similaires à celle de la figure 6.2.



FIGURE 6.2 – vaisseaux rétiniens et nervation alaire chez les insectes

PERSPECTIVES

L'étude de la feuille comme organe unique pour l'identification pourrait être étendue de deux manières :

– **1- Photographies de feuilles avec un fond naturel**

Nous avons considéré dans cette thèse des images de feuilles prises sur un fond uniforme. Une extension possible de notre approche serait d'associer une méthode de segmentation robuste pour pouvoir identifier l'espèce à partir des photographies de feuilles. Ceci permettra l'utilisation du système d'identification par un plus grand nombre d'utilisateurs.

– **2- Représentation plus complète de la feuille**

La micro-texture de la feuille n'est représentée dans notre travail que dans le scénario $eASC_{PS-PC}$ d'une façon locale dans le voisinage de chaque point saillant. Il serait judicieux de compléter la description existante de la feuille (contour, points saillants et nervures) par un descripteur global de texture. Une approche de type LBP (Local Binary Pattern) semble être prometteuse [WR08, RWZ12]. Ce nouveau descripteur pourrait être fusionné avec nos descripteurs.

De plus, de nouvelles stratégies de fusion précoce et tardive pourraient également être étudiées et comparées.

L'identification multi-organe et les données textuelles

Il est à noter que les botanistes utilisent aussi l'information fournie par d'autres organes par la requête multi-organe [GBB⁺12] quand la donnée de la feuille ne permet pas de déterminer l'espèce végétale. Dans le cadre du projet Pl@ntNet, des collectes de données ont été effectuées afin de construire une base qui comporte des images de différents organes d'arbres (feuille, fruit, fleur, écorce et rameau) [GJB⁺13].

Il serait judicieux de combiner les résultats obtenus sur les feuilles avec ceux obtenus sur d'autres organes pour conforter l'identification.

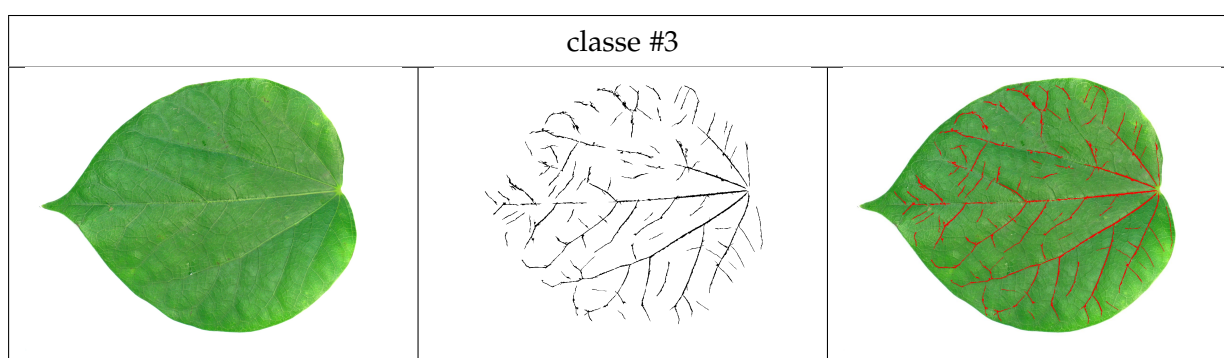
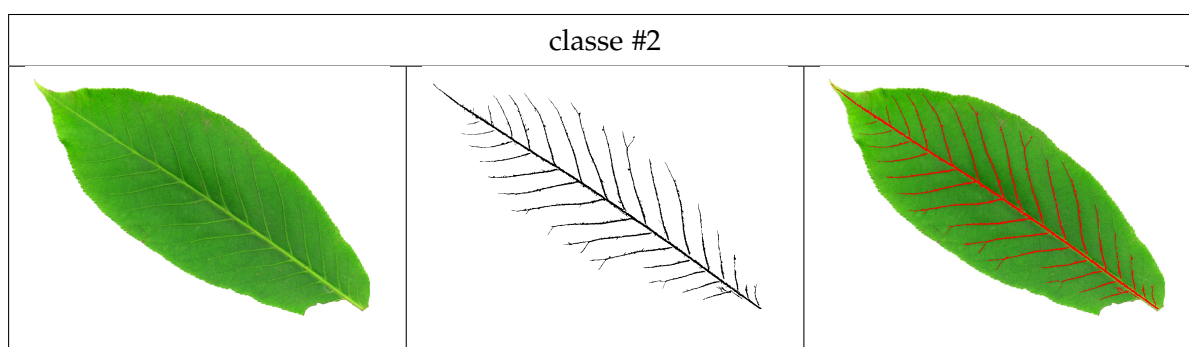
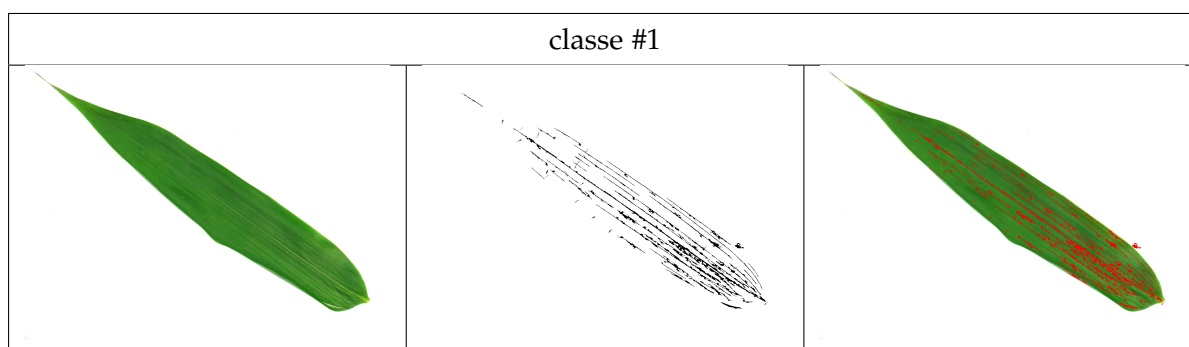
De plus, des connaissances statistiques sur la localisation des espèces par rapport aux données terrain (GPS, nature du sol, etc.) pourraient être prises en compte lors du processus d'identification.

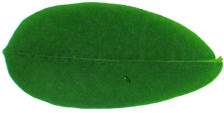

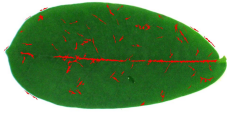

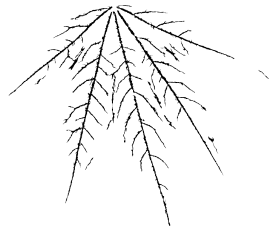


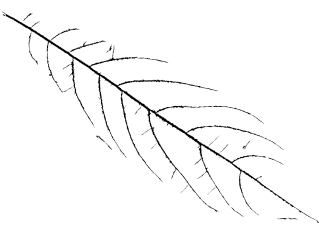
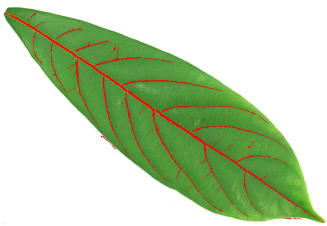

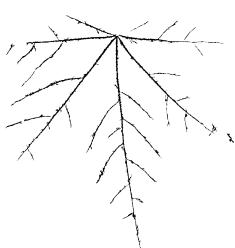

A

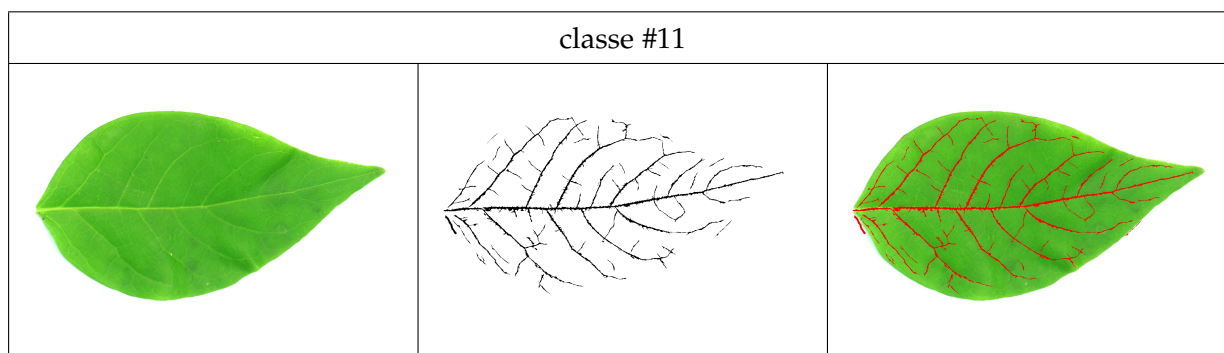
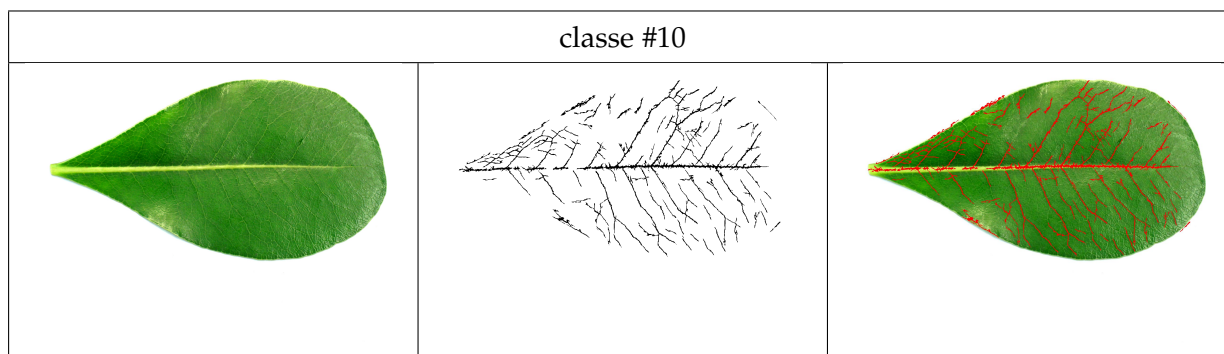
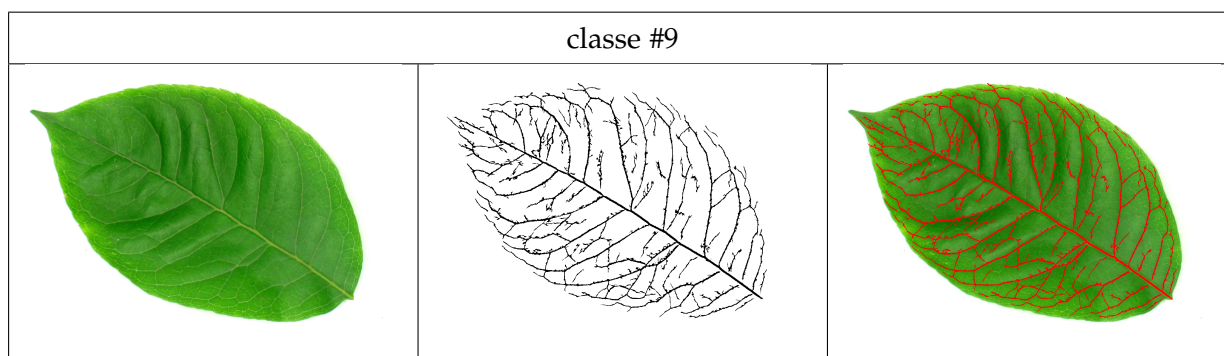
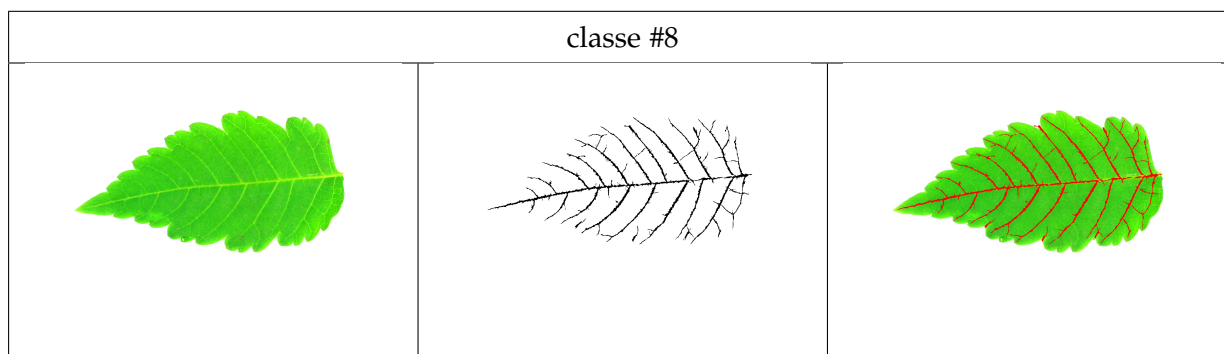
Résultats d'extraction de nervures sur la base Flavia

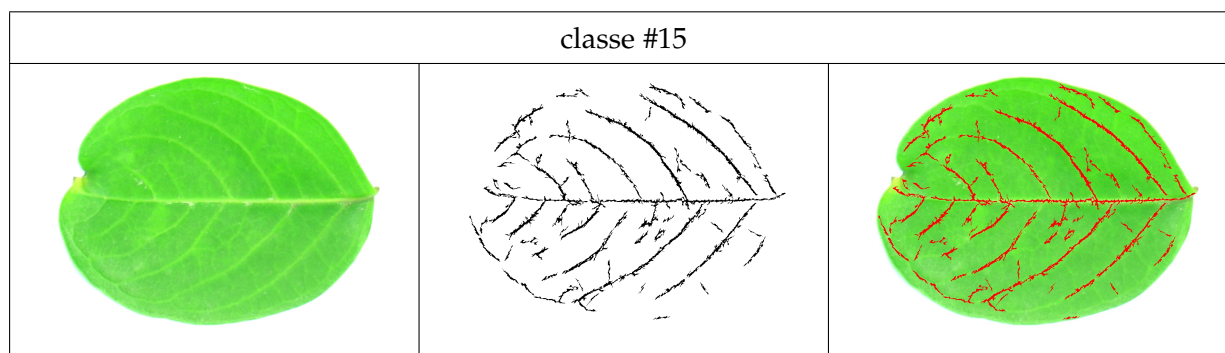
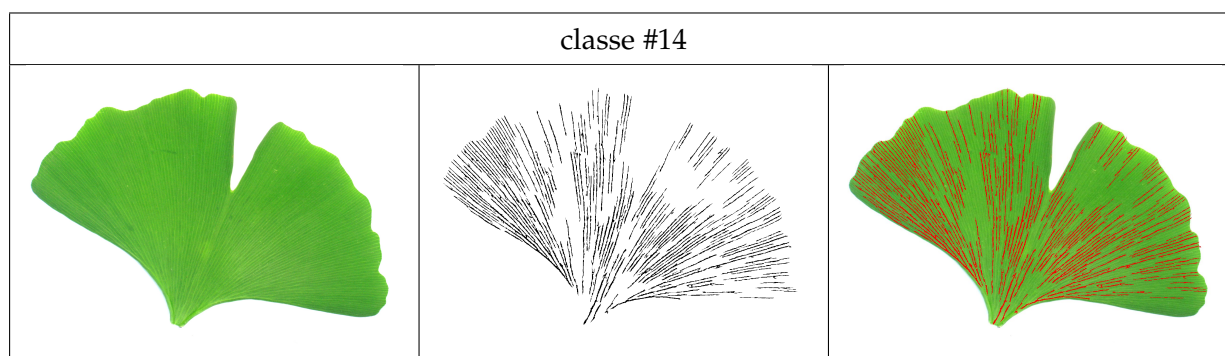
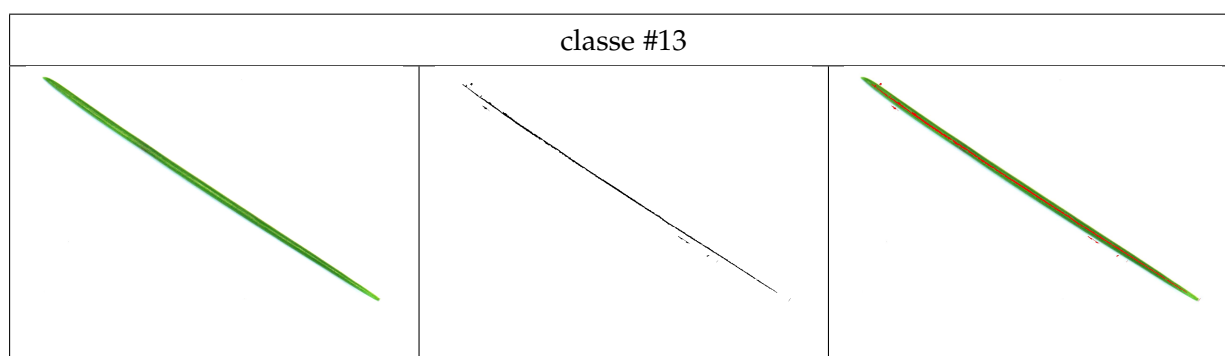
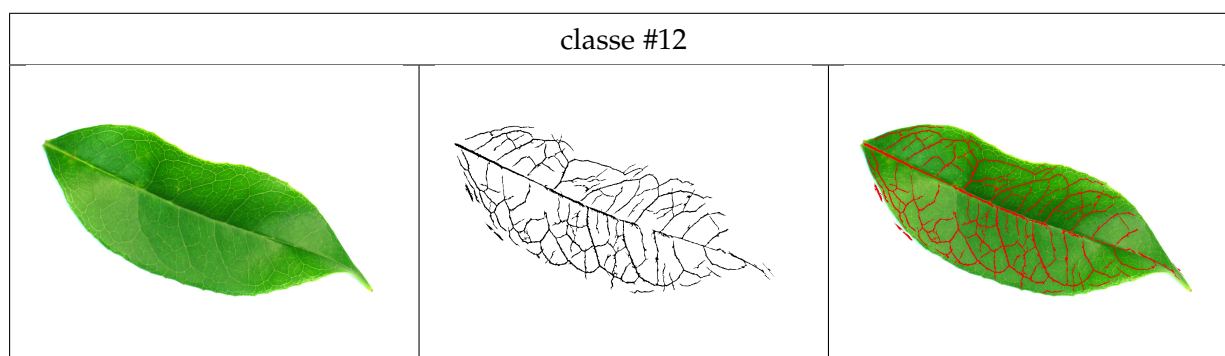
Nous consacrons cette annexe pour reporter les résultats d'extraction des nervures sur la base Flavia. Une image par classe est montrée.

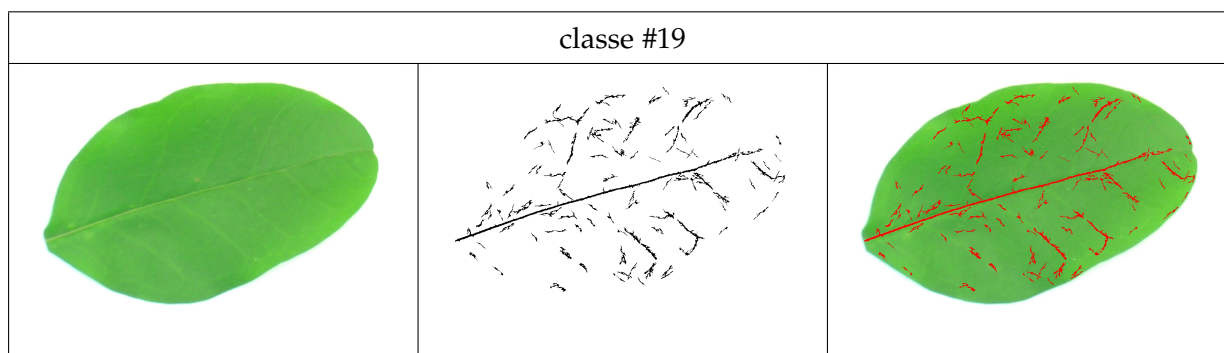
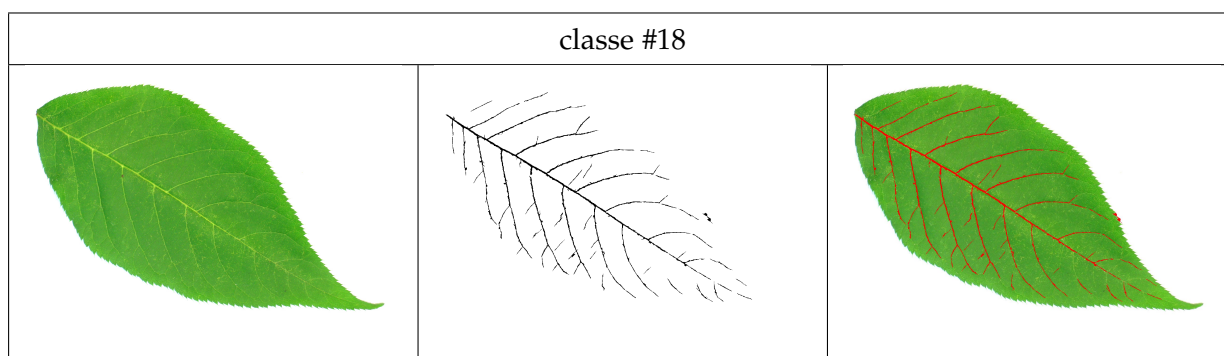
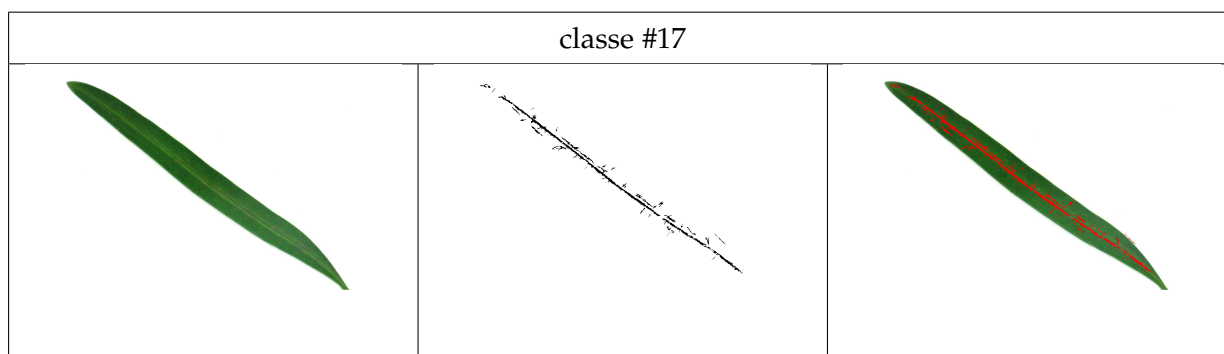
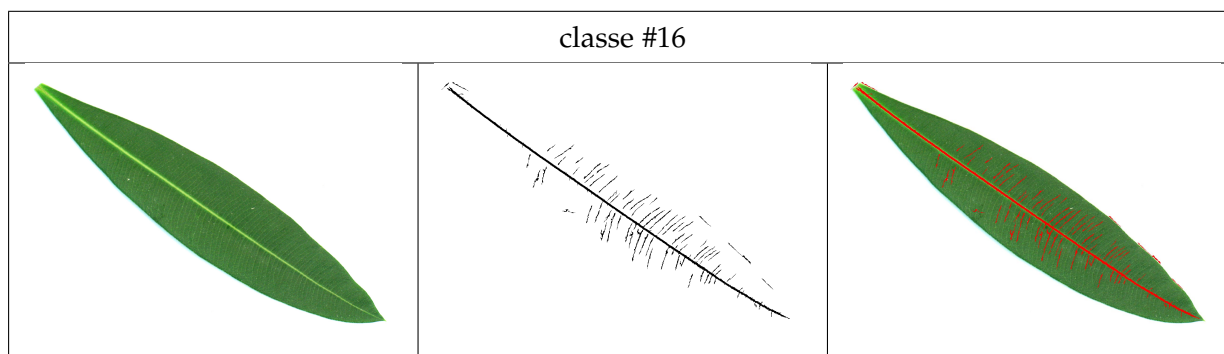
Les nervures principales et secondaires sont généralement extraites avec une bonne précision. Une partie des nervures tertiaires est également extraite pour certaines classes. Le nombre de faux positifs est relativement faible et dépend du contraste des nervures et de la qualité des images. Quand celle-ci est bonne, la précision d'extraction est élevée avec une absence quasi-totale du bruit, c'est le cas par exemple des images de la classe #24.

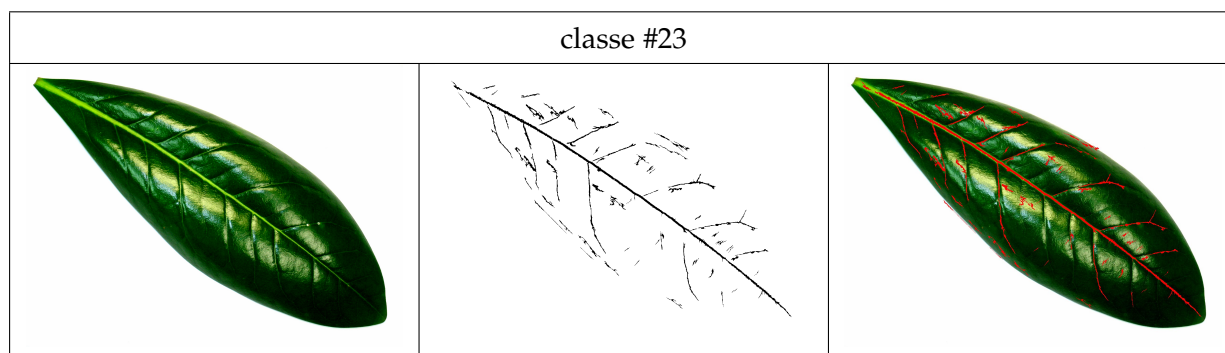
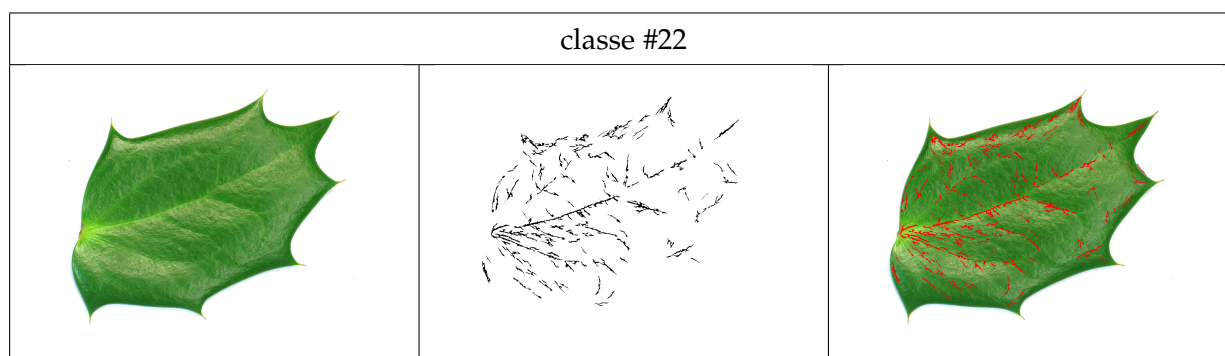
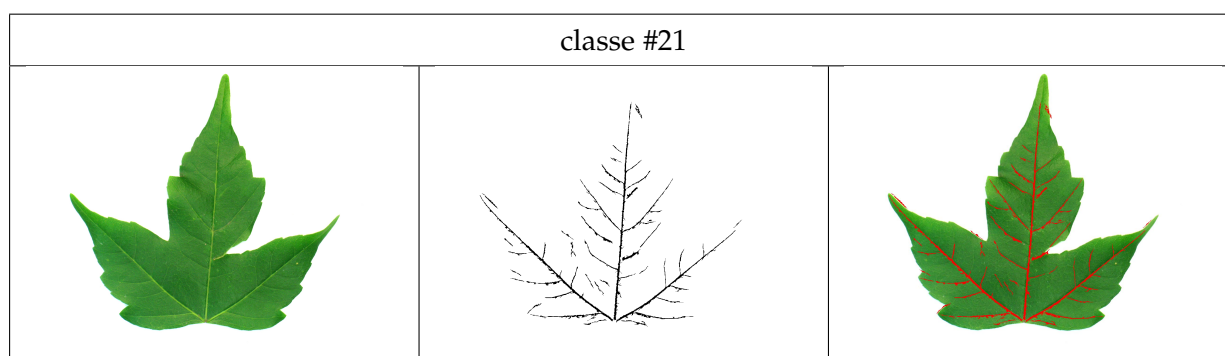
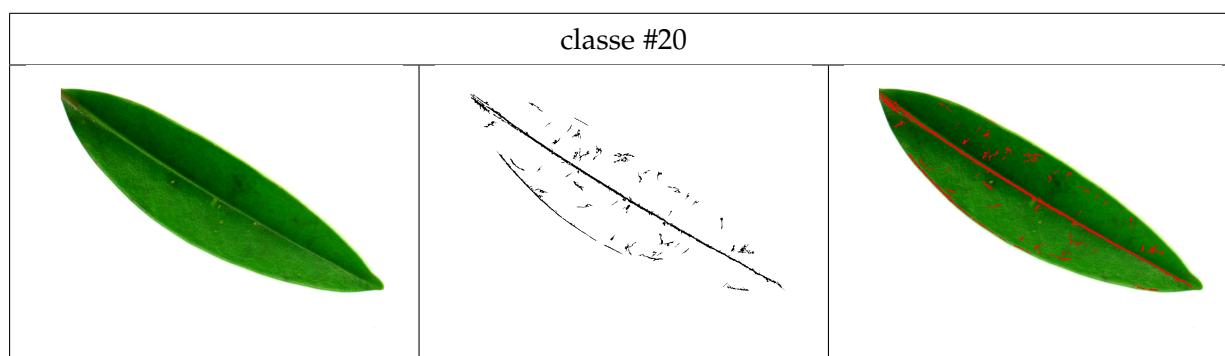


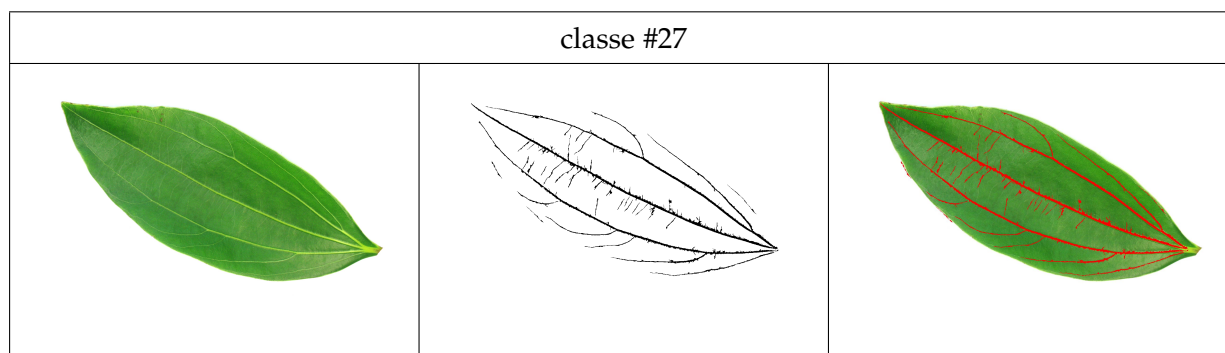
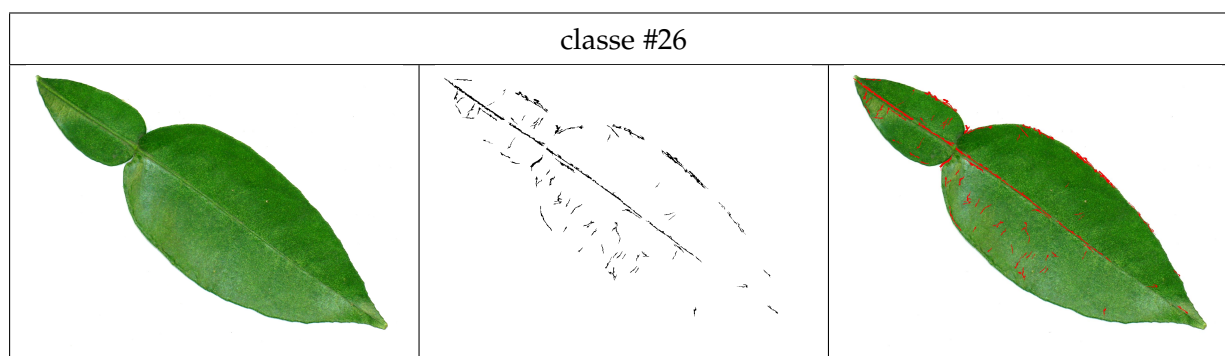
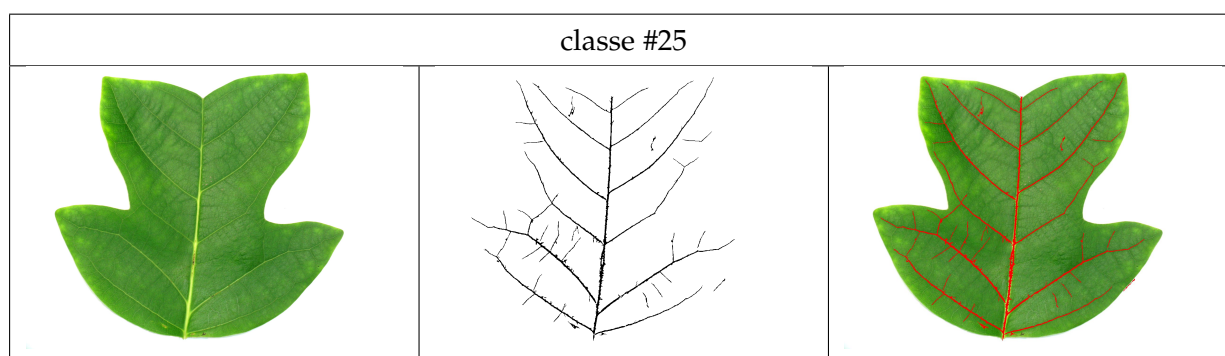
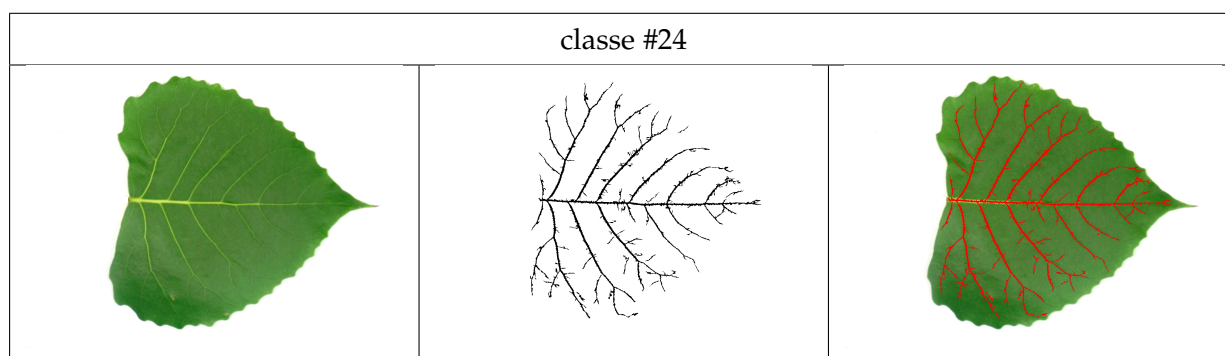
classe #4		
		
classe #5		
		
classe #6		
		
classe #7		
		

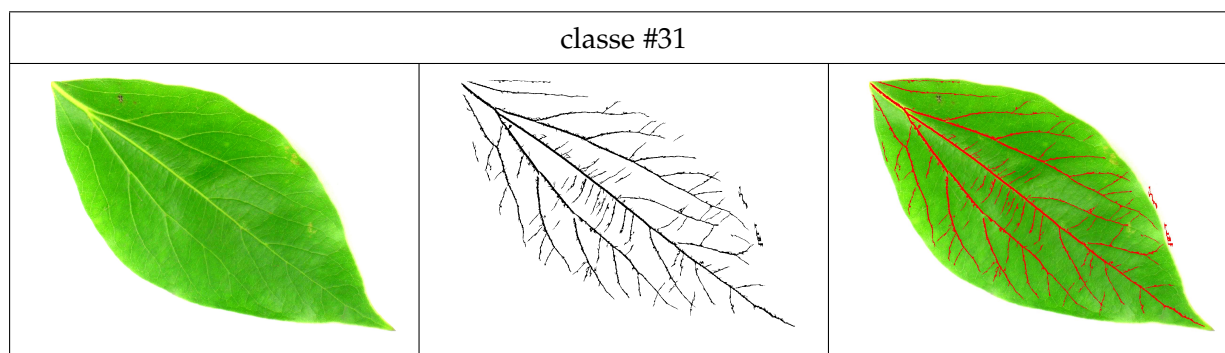
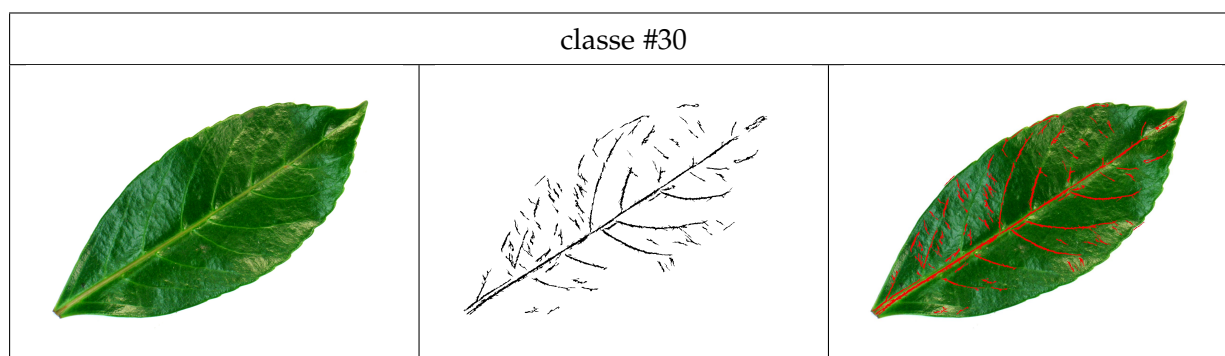
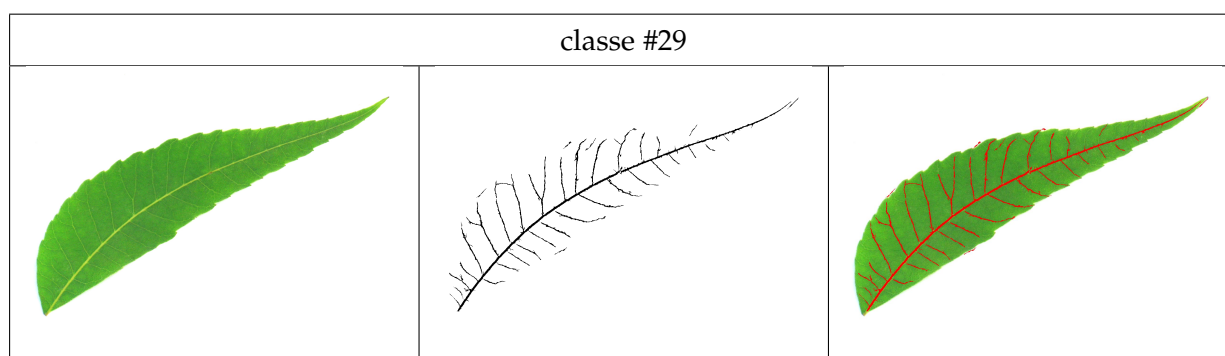
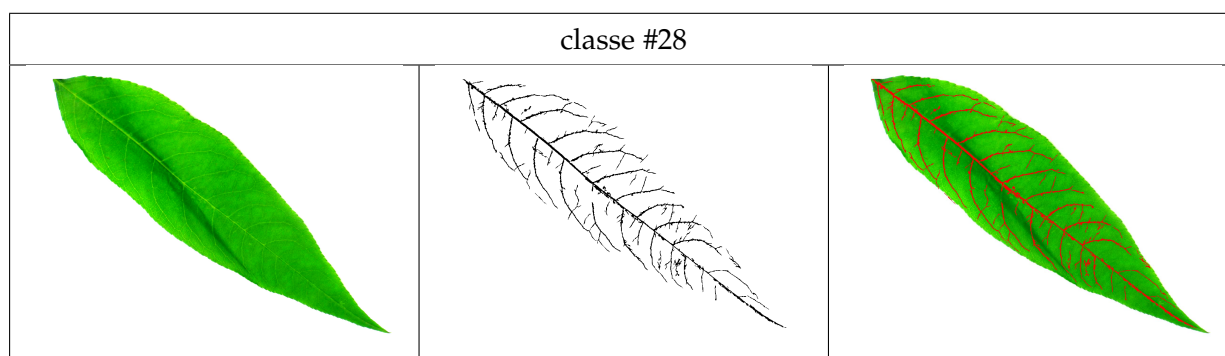















classe #32		
		

Bibliographie

- [AKF08] N. Alajlan, M.S. Kamel, and G.H. Freeman. Geometry-based image retrieval in binary image databases. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(6) :1003 –1013, jun 2008. (Cité pages 21 et 22.)
- [ARKF07] Naif Alajlan, Ibrahim El Rube, Mohamed S. Kamel, and George Freeman. Shape retrieval using triangle-area representation and dynamic space warping. *Pattern Recognition*, 40(7) :1911 – 1920, 2007. (Cité pages 21, 43, 50, 56, 103 et 104.)
- [ASF⁺11] Irshad Ahmad, Muhammad Hameed Siddiqi, Iram Fatima, Sungyoung Lee, and Young-Koo Lee. Weed classification based on haar wavelet transform via k-nearest neighbor (k-nn) for real-time automatic sprayer control system. In *Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication, ICUIMC '11*, pages 17 :1–17 :6, New York, NY, USA, 2011. ACM. (Cité page 37.)
- [ASFRRGG11] Juan Ignacio Arribas, Gonzalo V. Sánchez-Ferrero, Gonzalo Ruiz-Ruiz, and Jaime Gómez-Gil. Leaf classification in sunflower crops by computer vision and neural networks. *Computers and Electronics in Agriculture*, 78(1) :9–18, 2011. (Cité pages 15, 16 et 42.)
- [ASR07] J. Amores, N. Sebe, and P. Radeva. Context-based object-class recognition and retrieval by generalized correlograms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(10) :1818 –1833, oct 2007. (Cité page 66.)
- [ASS⁺11] A.M. Aibinu, M.-J.E. Salami, A. A. Shafie, N. Hazali, and N. Termidzi. Automatic fruits identification system using hybrid technique. In *Electronic Design, Test and Application (DELTA), 2011 Sixth IEEE International Symposium on*, pages 217–221, Jan 2011. (Cité page 10.)

- [BB09] André Backes and OdemirM. Bruno. Plant leaf identification using multi-scale fractal dimension. In Pasquale Foggia, Carlo Sansone, and Mario Vento, editors, *Image Analysis and Processing – ICIAP 2009*, volume 5716 of *Lecture Notes in Computer Science*, pages 143–150. Springer Berlin Heidelberg, 2009. (Cité pages 23, 28, 37 et 43.)
- [BB10] André Ricardo Backes and Odemir Martinez Bruno. Shape classification using complex network and multi-scale fractal dimension. *Pattern Recognition Letters*, 31(1) :44 – 51, 2010. (Cité pages 23 et 28.)
- [BCB09] André Ricardo Backes, Dalcimar Casanova, and Odemir Martinez Bruno. A complex network-based approach for boundary shape analysis. *Pattern Recognition*, 42(1) :54–67, 2009. (Cité pages 28 et 29.)
- [BCF⁺08] Peter Belhumeur, Daozheng Chen, Steven Feiner, David Jacobs, W. Kress, Haibin Ling, Ida Lopez, Ravi Ramamoorthi, Sameer Sheorey, Sean White, and Ling Zhang. Searching the world’s herbaria : A system for visual identification of plant species. In *European Conference on Computer Vision (ECCV)*, pages 116–129, 2008. (Cité pages 3 et 98.)
- [BdFd08] Odemir Martinez Bruno, Rodrigo de Oliveira Plotze, Mauricio Falvo, and Mário de Castro. Fractal dimension applied to plant identification. *Information Sciences*, 178(12) :2722 – 2733, 2008. (Cité pages 23, 32, 35 et 43.)
- [Bea78] P. R. Beaudet. Rotationally invariant image operators. In *Proceedings of the 4th International Joint Conference on Pattern Recognition*, pages 579–583, Kyoto, Japan, November 1978. (Cité page 63.)
- [BFF⁺01] N. Boujemaa, J. Fauqueur, M. Ferecatu, F. Fleuret, V. Gouet, B. Lesaux, and H. Sahbi. Ikona : Interactive specific and generic image retrieval. In *International workshop on Multimedia ContentBased Indexing and Retrieval (MMCBIR)*, 2001. (Cité pages 4, 5, 46, 47, 63 et 174.)
- [BLL⁺14] T. Berg, Jiongxin Liu, Seung Woo Lee, M.L. Alexander, D.W. Jacobs, and P.N. Belhumeur. Birdsnap : Large-scale fine-grained visual categorization of birds. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2019–2026, June 2014. (Cité page 2.)

- [BMP02] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4) :509 –522, apr 2002. (Cité pages 5, 24, 25, 66, 67 et 68.)
- [BPC⁺12] Branko Brkljac, Marko Panic, Dubravko Culibrk, Vladimir S. Crnojevic, Jelena Acanski, and Ante Vujic. Automatic hoverfly species discrimination. In *ICPRAM (2)*, pages 108–115, 2012. (Cité page 2.)
- [BYL⁺10] Xiang Bai, Xingwei Yang, L.J. Latecki, Wenyu Liu, and Zhuowen Tu. Learning context-sensitive shape similarity by graph transduction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(5) :861 –874, may 2010. (Cité page 108.)
- [BYM⁺12] Vera Bakic, Itheri Yahiaoui, Sofiène Mouine, Saloua Litayem Ouertani, Wajih Ouertani, Anne Verroust-Blondet, Hervé Goëau, and Alexis Joly. Inria IMEDIA2’s Participation at ImageCLEF 2012 Plant Identification Task. In *CLEF (Online Working Notes/Labs/Workshop) 2012*, Rome, Italy, September 2012. (Cité pages 143 et 176.)
- [CA10] Carlos Caballero and M. Carmen Aranda. Plant species identification using leaf image retrieval. In *ACM International Conference on Image and Video Retrieval (CIVR)*, pages 327–334, 2010. (Cité pages 15, 16, 18, 19, 21 et 43.)
- [CCC⁺12] James S. Cope, David Corney, Jonathan Y. Clark, Paolo Remagnino, and Paul Wilkin. Plant species identification using digital morphometrics : A review. *Expert Systems with Applications*, 39(8) :7562 – 7573, 2012. (Cité pages 10 et 31.)
- [CdMSJB09] Dalcimar Casanova, Jarbas Joaci de Mesquita Sá Junior, and Odemir Martinez Bruno. Plant leaf identification using gabor wavelets. *Int. J. Imaging Syst. Technol.*, 19(3) :236–243, September 2009. (Cité pages 36 et 43.)
- [Cer13] Guillaume Cerruti. *Segementation et interprétation d’images naturelles pour l’identification de feuilles d’arbres sur smartphone*. PhD thesis, Université Lumière Lyon 2, 2013. (Cité page 3.)
- [CFB11] Dalcimar Casanova, João Batista Florindo, and Odemir Martinez Bruno. IFSC/USP at ImageCLEF 2011 : Plant identification task. In *CLEF (Notebook Papers/Labs/Workshop)*, 2011. (Cité pages 28 et 29.)

- [CRBW10a] James Cope, Paolo Remagnino, Sarah Barman, and Paul Wilkin. The extraction of venation from leaf images by evolved vein classifiers and ant colony algorithms. In Jacques Blanc-Talon, Don Bone, Wilfried Philips, Dan Popescu, and Paul Scheunders, editors, *Advanced Concepts for Intelligent Vision Systems*, volume 6474 of *Lecture Notes in Computer Science*, pages 135–144. Springer Berlin Heidelberg, 2010. (Cité pages 33 et 34.)
- [CRBW10b] JamesS. Cope, Paolo Remagnino, Sarah Barman, and Paul Wilkin. Plant texture classification using gabor co-occurrences. In George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Ronald Chung, Riad Hammound, Muhammad Hussain, Tan Kar-Han, Roger Crawfis, Daniel Thalmann, David Kao, and Lisa Avila, editors, *Advances in Visual Computing*, volume 6454 of *Lecture Notes in Computer Science*, pages 669–677. Springer Berlin Heidelberg, 2010. (Cité pages 36 et 43.)
- [CTCV13] Guillaume Cerutti, Laure Tougne, Didier Coquin, and Antoine Vacavant. Curvature-scale-based contour understanding for leaf margin shape recognition and species identification. In *VISAPP*, February 2013. (Cité page 21.)
- [CTM⁺13] Guillaume Cerutti, Laure Tougne, Julien Mille, Antoine Vacavant, and Didier Coquin. Understanding leaves in natural images – a model-based approach for tree species identification. *Computer Vision and Image Understanding*, 117(10) :1482 – 1501, 2013. (Cité pages 3, 43, 45 et 50.)
- [DMR99] M. Das, R. Manmatha, and E.M. Riseman. Indexing flower patent images using domain knowledge. *Intelligent Systems and their Applications, IEEE*, 14(5) :24–33, Sep 1999. (Cité page 10.)
- [DWZ07] Ji-Xiang Du, Xiao-Feng Wang, and Guo-Jun Zhang. Leaf shape based plant species recognition. *Applied Mathematics and Computation*, 185(2) :883 – 893, 2007. (Cité pages 17, 18, 19, 28 et 43.)
- [EFZP00] M. S. El-Faki, N. Zhang, and D. E. Peterson. Weed detection using color machine vision. *ansactions of the ASAE*, 43 :1969–1978, 2000. (Cité page 38.)
- [Ell09] B. Ellis. *Manual of Leaf Architecture*. Cornell paperbacks. CABI, 2009. (Cité page 2.)

- [EM12] Raffi Enficiaud and Sofiène Mouine. Landmark extraction from leaves with palmate venation - application to grape. In *ICPRAM (2)*, pages 520–524, 2012. (Cité pages 32, 34 et 77.)
- [ERAK⁺05] I. El Rube, N. Alajlan, M. Kamel, M. Ahmed, and G. Freeman. Robust multiscale triangle-area representation for 2D shapes. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 1, pages I – 545–8, sep 2005. (Cité pages 16, 17, 21, 22 et 43.)
- [FC06] H. Fu and Z. Chi. Combined thresholding and neural network approach for vein pattern extraction from leaf images. *Vision, Image and Signal Processing, IEE Proceedings -*, 153(6) :881–892, 2006. (Cité pages 32 et 33.)
- [Fer05] Marin Ferecatu. *Image retrieval with active relevance feedback using both visual and keyword-based descriptors*. PhD thesis, University of Versailles Saint-Quentin-en-Yvelines, 2005. (Cité page 70.)
- [FGU91] E Franz, M.R. Gebhardt, and K.B. Unklesbay. The use of local spectral properties of leaves as an aid for identifying weed seedlings in digital images. *Transactions of the ASAE*, 34, 1991. (Cité pages 37, 38 et 43.)
- [FS07] Pedro F. Felzenszwalb and Joshua D. Schwartz. Hierarchical matching of deformable shapes. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8, 2007. (Cité page 108.)
- [FW08] Xiaofeng Fu and Wei Wei. Centralized binary patterns embedded with image euclidean distance or facial expression recognition. In *Natural Computation, 2008. ICNC '08. Fourth International Conference on*, volume 4, pages 115–119, 2008. (Cité page 36.)
- [GBB⁺12] Hervé Goëau, Pierre Bonnet, Julien Barbe, Vera Bakic, Alexis Joly, Jean-François Molino, Daniel Barthelemy, and Nozha Boujemaa. Multi-organ plant identification. In *Proceedings of the 1st ACM international workshop on Multimedia analysis for ecological data*, MAED '12, pages 41–44, New York, NY, USA, 2012. ACM. (Cité pages 10 et 177.)
- [GBJ⁺11] Hervé Goëau, Pierre Bonnet, Alexis Joly, Nozha Boujemaa, Daniel Barthelemy, Jean-François Molino, Philippe Birnbaum, Elise Mouysset, and Marie Picard. The CLEF 2011 plant images classification task. In *CLEF (Notebook Papers/Labs/Workshop)*, 2011. (Cité pages 6, 29, 44, 63, 131 et 162.)

- [GBJ⁺12] Hervé Goëau, Pierre Bonnet, Alexis Joly, Itheri Yahiaoui, Daniel Barthélémy, Nozha Boujemaa, and Jean-François Molino. The ImageCLEF 2012 Plant identification Task. In *CLEF 2012*, Rome, Italy, September 2012. (Cité pages 6, 140, 143 et 162.)
- [GBVC⁺13] Manuel Grand-Brochier, Antoine Vacavant, Guillaume Cerutti, Kevin Bianchi, and Laure Tougne. Comparative study of segmentation methods for tree leaves extraction. In *Proceedings of the International Workshop on Video and Image Ground Truth in Computer Vision Applications*, VIGTA '13, pages 7 :1–7 :7, New York, NY, USA, 2013. ACM. (Cité pages 42 et 50.)
- [GJB⁺13] Hervé Goëau, Alexis Joly, Pierre Bonnet, Vera Bakic, Daniel Barthélémy, Nozha Boujemaa, and Jean-François Molino. The imageclef plant identification task 2013. In *Proceedings of the 2Nd ACM International Workshop on Multimedia Analysis for Ecological Data*, MAED '13, pages 23–28, 2013. (Cité pages 6, 44, 162, 167, 168 et 177.)
- [GJY⁺11] Hervé Goëau, Alexis Joly, Itheri Yahiaoui, Pierre Bonnet, and Elise Mouysset. Participation of INRIA& Pl@ntNet to ImageCLEF 2011 plant images classification task. In *CLEF (Notebook Papers/Labs/Workshop)*, 2011. (Cité pages 39, 41 et 70.)
- [GN01] D.S. Guru and P. Nagabhushan. Triangular spatial relationship : a new approach for spatial knowledge representation. *Pattern Recognition Letters*, 22(9) :999 – 1006, 2001. (Cité page 51.)
- [Gou00] V. Gouet. *Mise en correspondance d'images en couleur*. PhD thesis, Université Montpellier II, 2000. (Cité page 63.)
- [GPRG12] A. Gopal, S. Prudhveeswar Reddy, and V. Gayatri. Classification of selected medicinal plants leaf using image processing. In *Machine Vision and Image Processing (MVIP), 2012 International Conference on*, pages 5–8, 2012. (Cité pages 16 et 17.)
- [GZZ10] Zhenhua Guo, D. Zhang, and D. Zhang. A completed modeling of local binary pattern operator for texture classification. *Image Processing, IEEE Transactions on*, 19(6) :1657–1663, June 2010. (Cité page 36.)

- [HCL⁺04] An-Xiang Hong, Gang Chen, Jun-Li Li, Zhe-Ru Chi, and Dan Zhang. A flower image retrieval method based on roi feature. *J Zhejiang Univ Sci*, 5(7) :764–72, 2004. (Cité page 10.)
- [HGBRM10] N.V. Hoàng, V. Gouet-Brunet, M. Rukoz, and M. Manouvrier. Embedding spatial information into image content description for scene retrieval. *Pattern Recognition*, 43(9) :3013 – 3024, 2010. (Cité pages 51 et 52.)
- [HHD⁺06] Zhi-Kai Huang, De-Shuang Huang, Ji-Xiang Du, Zhong-Hua Quan, and Shen-Bo Guo. Bark classification based on gabor filter features using rbpnn neural network. In Irwin King, Jun Wang, Lai-Wan Chan, and DeLiang Wang, editors, *Neural Information Processing*, volume 4233 of *Lecture Notes in Computer Science*, pages 80–87. Springer Berlin Heidelberg, 2006. (Cité page 10.)
- [HJSS92] Q. Huang, A.K. Jain, George C. Stockman, and A.J.M. Smucker. Automatic image analysis of plant root structures. In *Pattern Recognition, 1992. Vol.II. Conference B : Pattern Recognition Methodology and Systems, Proceedings., 11th IAPR International Conference on*, pages 569–572, Aug 1992. (Cité page 10.)
- [HMLZ14] Chao Huang, Fanman Meng, Wang Luo, and Shuyuan Zhu. Bird breed classification and annotation using saliency based graphical model. *Journal of Visual Communication and Image Representation*, 2(6) :1299 – 1307, 2014. (Cité page 2.)
- [HPS09] Marko Heikkilä, Matti Pietikäinen, and Cordelia Schmid. Description of interest regions with local binary patterns. *Pattern Recogn.*, 42(3) :425–436, March 2009. (Cité page 36.)
- [HS88] C. Harris and M. Stephens. A Combined Corner and Edge Detection. In *Proceedings of The Fourth Alvey Vision Conference*, pages 147–151, 1988. (Cité page 63.)
- [HT05] D. Frank Hsu and Isak Taksa. Comparing rank and score combination methods for data fusion in information retrieval. *Information Retrieval*, 8(3) :449–480, 2005. (Cité page 152.)
- [Hu62] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, 8(2) :179–187, 1962. (Cité page 27.)

- [IKM08] Dimitris K. Iakovidis, Eystratios G. Keramidas, and Dimitris Maroulis. Fuzzy local binary patterns for ultrasound texture characterization. In Aurélio Campilho and Mohamed Kamel, editors, *Image Analysis and Recognition*, volume 5112 of *Lecture Notes in Computer Science*, pages 750–759. 2008. (Cité page 36.)
- [JB08] Alexis Joly and Olivier Buisson. A posteriori multi-probe locality hashing. In *16th ACM international conference on Multimedia*, pages 209–218, 2008. (Cité pages 39 et 47.)
- [JB09] Alexis Joly and Olivier Buisson. Logo retrieval with a contrario visual query expansion. In *17th ACM international conference on Multimedia*, pages 581–584, 2009. (Cité page 39.)
- [JB11] Alexis Joly and Olivier Buisson. Random maximum margin hashing. In *CVPR*, pages 873–880, 2011. (Cité page 39.)
- [JHDH06] Mladen Jović, Yutaka Hatakeyama, Fangyan Dong, and Kaoru Hirota. Image retrieval based on similarity score fusion from feature similarity ranking lists. In *Third Int. Conf. on Fuzzy Systems and Knowledge Discovery (FSKD 2006)*, pages 461–470, Berlin, Heidelberg, 2006. Springer-Verlag. (Cité page 153.)
- [KBB⁺12] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. B. Soares. Leafsnap : A computer vision system for automatic plant species identification. In *12th European Conference on Computer Vision (ECCV 2012)*, pages 502–516, Florence, Italy, October 2012. (Cité pages 3, 21 et 43.)
- [KG82] Frank P Kuhl and Charles R Giardina. Elliptic fourier features of a closed contour. *Computer Graphics and Image Processing*, 18(3) :236 – 258, 1982. (Cité page 29.)
- [KK14] Yılmaz Kaya and Lokman Kayci. Application of artificial neural network for automatic detection of butterfly species using color and texture features. *The Visual Computer*, 30(1) :71–79, 2014. (Cité page 2.)
- [KNSS11a] A. Kadir, L. E. Nugroho, A. Susanto, and P.I. Santosa. Neural network application on foliage plant identification. *International Journal of Computer Applications*, 29(9) :15–22, sep 2011. (Cité pages 16, 42 et 43.)

- [KNSS11b] Abdul Kadir, Lukito Edi Nugroho, Adhi Susanto, and Paulus Insap Santosa. Foliage plant retrieval using polar fourier transform, color moments and vein features. *CoRR*, abs/1110.1513, 2011. (Cité pages 29, 30, 32, 33, 34, 38, 71, 89, 90, 110, 111, 157, 158, 170 et 176.)
- [KR82] L. Kitchen and A. Rosenfeld. Gray-level corner detection. *Pattern Recognition Letters*, 1(2) :95 – 102, 1982. (Cité page 63.)
- [KSL14] Sebastian Kurtek, Mo Shen, and Hamid Laga. Elastic reflection symmetry based shape descriptors. In *IEEE Winter Conference on Applications of Computer Vision, Steamboat Springs, CO, USA, March 24-26, 2014*, pages 293–300, 2014. (Cité pages 110, 170, 171 et 176.)
- [KYU11] Hanife Kebapci, Berrin Yanikoglu, and Gozde Unal. Plant image retrieval using color, shape and texture features. *Comput. J.*, 54(9) :1475–1490, September 2011. (Cité pages 36, 37, 38 et 43.)
- [Lee97] Joon Ho Lee. Analyses of multiple evidence combination. In *20th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '97*, pages 267–276, New York, NY, USA, 1997. ACM. (Cité page 153.)
- [LJ07] Haibin Ling and D.W. Jacobs. Shape classification using the inner-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2) :286 –299, February 2007. (Cité pages 24, 25, 43, 66, 101, 106 et 107.)
- [LKJB12] Jiongxin Liu, Angjoo Kanazawa, David Jacobs, and Peter Belhumeur. Dog breed classification using part localization. In *Computer Vision – ECCV 2012*, volume 7572 of *Lecture Notes in Computer Science*, pages 172–185. Springer Berlin Heidelberg, 2012. (Cité page 2.)
- [LKS⁺12] H. Laga, S. Kurtek, A. Srivastava, M. Golzarian, and Stanley Miklavcic. A riemannian elastic metric for shape-based plant leaf classification. In *Digital Image Computing : Techniques and Applications*, 2012. (Cité pages 110 et 126.)
- [LLE00] Longin Jan Latecki, Rolf Lakämper, and Ulrich Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 424–429, 2000. (Cité pages 22, 25 et 104.)

- [LW12] Hongzhi Liu and D. Frank Wu, Zhonghai and Hsu. Combination of multiple retrieval systems using rank-score function and cognitive diversity. In *AINA*, pages 167–174, 2012. (Cité page 152.)
- [LZWM08] Feng-Yan Lin, Chun-Hou Zheng, Xiao-Feng Wang, and Qing-Kui Man. Multiple classification of plant leaves based on gabor transform and lbp operator. In De-Shuang Huang, Donald C. Wunsch II, Daniel S. Levine, and Kang-Hyun Jo, editors, *ICIC (3)*, volume 15 of *Communications in Computer and Information Science*, pages 432–439, 2008. (Cité pages 36 et 43.)
- [MA04] F. Mokhtarian and S. Abbasi. Matching shapes with self-intersections : application to leaf classification. *IEEE Transactions on Image Processing*, 13, May 2004. (Cité pages 17, 21 et 62.)
- [MAK96] Farzin Mokhtarian, Sadegh Abbasi, and Josef Kittler. Robust and efficient shape indexing through curvature scale space. In *British Machine Vision Conference (BMVC)*, 1996. (Cité page 20.)
- [MCGB13] Bruno Brandoli Machado, Dalcimar Casanova, Wesley Nunes Gonçalves, and Odemir Martinez Bruno. Partial differential equations and fractal analysis to plant leaf identification. *Journal of Physics : Conference Series*, 410(1) :012066, 2013. (Cité pages 23 et 37.)
- [MCO13] Charles Mallah, James Cope, and James Orwell. Plant leaf classification using probabilistic integration of shape, texture and margin features. In *Signal Processing, Pattern Recognition and Applications*, 2013. (Cité pages 6, 98 et 127.)
- [MKJ08] Yang Mingqiang, Kpalma Kidiyo, and Ronsin Joseph. A survey of shape feature extraction techniques. *Pattern Recognition Techniques, Technology and Applications*, 2008. ISBN : 978-953-7619-24-4, InTech. (Cité page 13.)
- [Mor77] Hans P. Moravec. Towards automatic visual obstacle avoidance. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'77*, pages 584–584, San Francisco, CA, USA, 1977. Morgan Kaufmann Publishers Inc. (Cité page 62.)
- [MS89] F. Meyer and J. Serra. Contrasts and activity lattice. *Signal Processing*, 16(4) :303 – 317, 1989. (Cité page 78.)

- [MTA⁺11] Camilo Mora, Derek P. Tittensor, Sina Adl, Alastair G. B. Simpson, and Boris Worm. How Many Species Are There on Earth and in the Ocean? *PLoS Biol*, 9(8) :e1001127+, August 2011. (Cité page 2.)
- [MYB12] Olfa Mzoughi, Itheri Yahiaoui, and Nozha Boujemaa. Alignment of 2d objects for shape interpretation. In *13th International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2012, Dublin, Ireland, May 23-25, 2012*, pages 1–4, 2012. (Cité page 17.)
- [MYVB12] Sofiene Mouine, Itheri Yahiaoui, and Anne Verroust-Blondet. Advanced shape context for plant species identification using leaf image retrieval. In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, pages 49 :1–49 :8, 2012. (Cité page 41.)
- [MYVB13] Sofiène Mouine, Itheri Yahiaoui, and Anne Verroust-Blondet. A shape-based approach for leaf classification using multiscaletriangular representation. In *ICMR*, pages 127–134, 2013. (Cité page 101.)
- [NHK05] Yunyoung Nam, Eenjun Hwang, and Dongyoon Kim. Clover : A mobile content-based leaf image retrieval system. In EdwardA. Fox, ErichJ. Neuhold, Pimrumpai Premssmit, and Vilas Wuwongse, editors, *Digital Libraries : Implementing Strategies and Sharing Experiences*, volume 3815 of *Lecture Notes in Computer Science*, pages 139–148. Springer Berlin Heidelberg, 2005. (Cité page 3.)
- [NHK08] Yunyoung Nam, Eenjun Hwang, and Dongyoon Kim. A similarity-based leaf image retrieval scheme : Joining shape and venation features. *Computer Vision and Image Understanding*, 110(2) :245 – 259, 2008. (Cité pages 33, 34, 43, 62, 71 et 94.)
- [Nil09] M-E. Nilsback. *An Automatic Visual Flora – Segmentation and Classification of Flowers Images*. PhD thesis, University of Oxford, 2009. (Cité page 10.)
- [NLB12] Loris Nanni, Alessandra Lumini, and Sheryl Brahnam. Survey on lbp based texture descriptors for image classification. *Expert Systems with Applications*, 39(3) :3634 – 3641, 2012. (Cité pages 36 et 43.)
- [NMJS06] Joao Camargo Neto, George E. Meyer, David D. Jones, and Ashok K. Samal. Plant species identification using elliptic Fourier leaf shape analysis. *Com-*

- puters and Electronics in Agriculture*, 50(2) :121 – 134, 2006. (Cité pages 29 et 30.)
- [NZ08] M-E. Nilsback and A. Zisserman. Automated flower classification over a large number of classes. In *Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing*, Dec 2008. (Cité pages 10, 42 et 176.)
- [ON00] Mari Oide and Seishi Ninomiya. Discrimination of soybean leaflet shape by neural networks with image input. *Computers and Electronics in Agriculture*, 29(1-2) :59 – 72, 2000. (Cité page 42.)
- [OPH96] Timo Ojala, Matti Pietikäinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1) :51 – 59, 1996. (Cité page 35.)
- [OPM02] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7) :971–987, 2002. (Cité page 35.)
- [Ots79] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1) :62–66, 1979. (Cité pages 45, 46 et 50.)
- [PG08] P. Punitha and D.S. Guru. Symbolic image indexing and retrieval by spatial similarity : An approach based on b-tree. *Pattern Recognition*, 41(6) :2068 – 2085, 2008. (Cité page 51.)
- [PHN06] Jin-Kyu Park, EenJun Hwang, and Yunyoung Nam. A venation-based leaf image classification scheme. In HweeTou Ng, Mun-Kew Leong, Min-Yen Kan, and Donghong Ji, editors, *Information Retrieval Technology*, volume 4182 of *Lecture Notes in Computer Science*, pages 416–428. Springer Berlin Heidelberg, 2006. (Cité pages 31, 62 et 94.)
- [PHN08] JinKyu Park, EenJun Hwang, and Yunyoung Nam. Utilizing venation features for efficient leaf image retrieval. *Journal of Systems and Software*, 81(1) :71 – 82, 2008. (Cité pages 31, 33, 34, 62, 71 et 94.)
- [RnL10] Marçal Rusiñol and Josep Lladós. Efficient logo retrieval through hashing shape context descriptors. In *Proceedings of the 9th IAPR International Work-*

- shop on Document Analysis Systems, DAS '10*, pages 215–222, 2010. (Cité page 68.)
- [RS03] M. Elena Renda and Umberto Straccia. Web metasearch : rank vs. score based rank aggregation methods. In *ACM symposium on Applied computing, SAC '03*, pages 841–846, New York, NY, USA, 2003. ACM. (Cité page 152.)
- [RSBG13] Asma Rejeb Sfar, Nozha Boujemaa, and Donald Geman. Identification of plants from multiple images and botanical idkeys. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval, ICMR '13*, pages 191–198, New York, NY, USA, 2013. (Cité pages 17, 42 et 176.)
- [RSH96] Yong Rui, Alfred C. She, and Thomas S. Huang. Modified fourier descriptors for shape representation – a practical approach. In *PROC OF FIRST INTERNATIONAL WORKSHOP ON IMAGE DATABASES AND MULTIMEDIA SEARCH*, 1996. (Cité page 29.)
- [RWZ12] Xiao-Ming Ren, Xiao-Feng Wang, and Yang Zhao. An efficient multi-scale overlapped block lbp approach for leaf image recognition. In *Proceedings of the 8th international conference on Intelligent Computing - Theories and Applications, ICIC'12*, pages 237–243, Berlin, Heidelberg, 2012. Springer-Verlag. (Cité pages 42, 98 et 177.)
- [SCTK98] Daniel Sharvit, Jacky Chan, Hüseyin Tek, and Benjamin B. Kimia. Symmetry-based indexing of image databases. *J. VISUAL COMMUNICATION AND IMAGE REPRESENTATION*, 9 :366–380, 1998. (Cité pages 25 et 104.)
- [SGDS⁺10] Concetto Spampinato, Daniela Giordano, Roberto Di Salvo, Yun-Heh Jessica Chen-Burger, Robert Bob Fisher, and Gayathri Nadarajan. Automatic fish classification for underwater species behavior understanding. In *Proceedings of the First ACM International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams, ARTEMIS '10*, pages 45–50, 2010. (Cité page 2.)
- [SGG10] Krishna Singh, Indra Gupta, and Sangeeta Gupta. Svm-bdt pnn and fourier moment technique for classification of leaf shape. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, pages 67–78, 2010. (Cité page 32.)

- [SI07] E. Shechtman and M. Irani. Matching local self-similarities across images and videos. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, june 2007. (Cité page 66.)
- [SJ13] JoãoV.B. Soares and DavidW. Jacobs. Efficient segmentation of leaves in semi-controlled conditions. *Machine Vision and Applications*, 24(8) :1623–1643, 2013. (Cité pages 45 et 50.)
- [SK00] T. Saitoh and T. Kaneko. Automatic recognition of wild flowers. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, pages 507–510 vol.2, 2000. (Cité pages 17 et 42.)
- [SKK04] Thomas B. Sebastian, Philip N. Klein, and Benjamin B. Kimia. Recognition of shapes by editing their shock graphs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(5) :550–571, May 2004. (Cité pages 25 et 104.)
- [SLW12] KC Santosh, B. Lamiroy, and L.t Wendling. Symbol recognition using spatial relations. *Pattern Recognition Letters*, 33(3) :331–341, 2012. (Cité page 52.)
- [SMB98] C. Schmid, R. Mohr, and C. Bauckhage. Comparing and evaluating interest points. In *Computer Vision, 1998. Sixth INTERNATIONAL CONFERENCE on*, pages 230–235, Jan 1998. (Cité page 63.)
- [SMB00] Cordelia Schmid, Roger Mohr, and Christian Bauckhage. Evaluation of Interest Point Detectors. *International Journal of Computer Vision*, 37(2) :151–172, 2000. (Cité page 63.)
- [Söd01] Oskar J O Söderkvist. Computer vision classification of leaves from swedish trees. Master’s thesis, Linköping University, SE-581 83 Linköping, Sweden, September 2001. (Cité pages 6, 19, 24, 26, 98 et 108.)
- [SW11] Xin Shu and Xiao-Jun Wu. A novel contour descriptor for 2d shape matching and its application to image retrieval. *Image and Vision Computing*, 29(4) :286 – 294, 2011. (Cité page 25.)
- [SWL14] K. C. Santosh, L. Wendling, and B. Lamiroy. BoR : Bag-of-relations for symbol retrieval. *International Journal of Pattern Recognition and Artificial Intelligence*, 28(06) :1450017, 2014. (Cité page 52.)
- [Tea80] Michael Reed Teague. Image analysis via the general theory of moments. *J. Opt. Soc. Am.*, 70(8) :920–930, Aug 1980. (Cité page 27.)

- [TFdFC04] R.daS. Torres, A.X. Falcão, and L. da F. Costa. A graph-based approach for multiscale shape analysis. *Pattern Recognition*, 37(6) :1163 – 1174, 2004. (Cité page 23.)
- [UU12] C. Uluturk and A. Ugur. Recognition of leaves based on morphological features derived from two half-regions. In *Innovations in Intelligent Systems and Applications (INISTA), 2012 International Symposium on*, pages 1 –4, july 2012. (Cité pages 19, 42, 43, 110 et 111.)
- [Wal03] Thomas Walter. Application de la morphologie mathématique au diagnostic de la rétinopathie diabétique à partir d’images couleur. Master’s thesis, Ecole des Mines, Paris, 2003. (Cité pages 78 et 176.)
- [WBX⁺07] S.G. Wu, F.S. Bao, E.Y. Xu, Yu-Xuan Wang, Yi-Fan Chang, and Qiao-Liang Xiang. A leaf recognition algorithm for plant classification using probabilistic neural network. In *Signal Processing and Information Technology, 2007 IEEE International Symposium on*, pages 11 –16, dec. 2007. (Cité pages 6, 15, 16, 17, 20, 42, 79, 98 et 109.)
- [WCF03] Z. Wang, Z. Chi, and D. Feng. Shape based leaf image retrieval. *IEE Proceedings on Vision, Image and Signal Processing*, 150(1) :34 – 43, February 2003. (Cité pages 18 et 43.)
- [WCFW00] Zhiyong Wang, Zheru Chi, Dagan Feng, and Qing Wang. Leaf image retrieval with shape features. In Robert Laurini, editor, *Advances in Visual Information Systems*, volume 1929 of *Lecture Notes in Computer Science*, pages 477–487. Springer Berlin Heidelberg, 2000. (Cité pages 18 et 23.)
- [WFB⁺14] Xinggang Wang, Bin Feng, Xiang Bai, Wenyu Liu, and Longin Jan Latecki. Bag of contour fragments for robust shape classification. *Pattern Recognition*, (0) :-, 2014. (Cité pages 25, 26, 27 et 98.)
- [WG13] Bin Wang and Yongsheng Gao. Fast and effective retrieval of plant leaf shapes. In KyoungMu Lee, Yasuyuki Matsushita, JamesM. Rehg, and Zha-nyi Hu, editors, *Computer Vision – ACCV 2012*, volume 7725 of *Lecture Notes in Computer Science*, pages 475–486. Springer Berlin Heidelberg, 2013. (Cité pages 19, 43, 98, 103 et 104.)
- [WHD⁺08] Xiao-Feng Wang, De-Shuang Huang, Ji-Xiang Du, Huan Xu, and Laurent Heutte. Classification of plant leaf images with complicated back-

- ground. *Applied Mathematics and Computation*, 205(2) :916 – 926, 2008. <ce :title>Special Issue on Advanced Intelligent Computing Theory and Methodology in Applied Mathematics and Computation</ce :title>. (Cité page 28.)
- [WHT08] Lior Wolf, Tal Hassner, and Yaniv Taigman. Descriptor based methods in the wild. In *In : Faces in Real-Life Images Workshop in ECCV. (2008) (b) Similarity Scores based on Background Samples*, 2008. (Cité page 36.)
- [WMVBM95] D.M Woebbecke, G.E Meyer, K. Von Bargen, and A.D Mortensen. Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the ASAE*, 38(1) :259–269, 1995. (Cité pages 37 et 38.)
- [WPDP12] J. Bastow Wilson, Robert K. Peet, Jürgen Dengler, and Meelis Pärtel. Plant species richness : the world records. *Journal of Vegetation Science*, 23(4) :796–802, 2012. (Cité page 3.)
- [WR08] Jianixn Wu and J.M. Rehg. Where am i : Place instance and category recognition using spatial pact. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1 –8, jun 2008. (Cité pages 108, 176 et 177.)
- [XHS08] Jun Xie, Pheng-Ann Heng, and Mubarak Shah. Shape matching and modeling using skeletal context. *Pattern Recognition*, 41(5) :1756 – 1767, 2008. (Cité pages 25 et 26.)
- [YAT12] Berrin Yanikoglu, Erchan Aptoula, and Caglar Tirkaz. Sabanci-Okan system at ImageClef 2012 : Combining features and classifiers for plant identification. In *CLEF (Online Working Notes/Labs/Workshop)*, 2012. (Cité pages 38 et 176.)
- [YAY13] Berrin A. Yanikoglu, Erchan Aptoula, and S. Tolga Yildiran. Sabanci-okan system at imageclef 2013 plant identification competition. In *Working Notes for CLEF 2013 Conference , Valencia, Spain, September 23-26, 2013.*, 2013. (Cité page 176.)
- [YION04] Yosuke Yoshioka, Hiroyoshi Iwata, Ryo Ohsawa, and Seishi Ninomiya. Analysis of petal shape variation of *primula sieboldii* by elliptic fourier descriptors and principal component analysis. *Annals of Botany*, 94(5) :657–664, 2004. (Cité page 30.)

- [YMB12] Itheri Yahiaoui, Olfa Mzoughi, and Nozha Boujemaa. Leaf shape descriptor for tree species identification. In *Proceedings of the 2012 IEEE International Conference on Multimedia and Expo, ICME 2012*, pages 254–259, 2012. (Cité pages 16, 17, 138, 140 et 176.)
- [ZGZL10] Baochang Zhang, Yongsheng Gao, Sanqiang Zhao, and Jianzhuang Liu. Local derivative pattern versus local binary pattern : Face recognition with high-order local pattern descriptor. *Image Processing, IEEE Transactions on*, 19(2) :533–544, Feb 2010. (Cité page 36.)
- [Zhe00] Miao Zhenjiang. Zernike moment-based image shape analysis and its application. *Pattern Recognition Letters*, 21(2) :169 – 177, 2000. (Cité page 28.)
- [ZHW12] Zhi-De Zhi, Rong-Xiang Hu, and Xiao-Feng Wang. A new weighted arc-sc approach for leaf image recognition. In *Intelligent Computing Theories and Applications*, volume 7390 of *Lecture Notes in Computer Science*, pages 503–509. 2012. (Cité pages 24, 25 et 98.)
- [ZSM11] Z. Zulkifli, Puteh Saad, and I.A. Mohtar. Plant leaf identification using moment invariants and general regression neural network. In *Hybrid Intelligent Systems (HIS), 2011 11th International Conference on*, pages 430–435, 2011. (Cité pages 28 et 42.)

Identification d'espèces végétales par une description géométrique locale d'images de feuilles

Sofiene MOUINE

RESUME : L'identification d'une plante se base généralement sur l'observation de ses organes. Nous avons utilisé l'analyse d'images pour élaborer un processus de reconnaissance automatique de l'espèce végétale à partir de la donnée d'une image de feuille. La forme de la feuille est représentée par des descripteurs triangulaires multi-échelle locaux associés aux points échantillonnés sur le contour. Une représentation dérivée du contexte de forme (Shape Context) décrit ses points saillants. La nervation est extraite avec des outils morphologiques et le réseau de nervures est décrit par ses directions principales et sa répartition spatiale dans le contexte de la surface de la feuille.

Les évaluations montrent que nos approches permettent une bonne d'identification de l'espèce végétale et que la fusion tardive des descripteurs augmente sa précision.

MOTS-CLEFS : identification des espèces végétales, descripteur de forme 2D, représentation triangulaire multi-échelle, contexte de forme, mise en correspondance locale, fusion tardive.

ABSTRACT : Plant species identification is based on the observation of its organs and mostly on visual criteria. This enables the use of computer vision and pattern recognition techniques for developing an automatic recognition process of the plant species from a leaf image. The shape of the leaf is represented by local descriptors based on multi-scale triangular representations associated to a set of points sampled on the contour. A shape context based representation describes its saillant points. The venation is extracted by detecting elementary linear structures with morphological tools and described by its main directions and its spatial distribution in the context of the leaf boundary. Evaluations show that our approaches identify the plant species of the leaf in most of the cases and that the late fusion of the proposed descriptors improves the identification process.

KEY-WORDS : plant species identification, 2D shape descriptor, multi-scale triangular representation, shape context, local features matching, late fusion.